Spring 7-6-2018

Comparative Genetic and Genomic Analysis of the Novel Fusellovirus Sulfolobus Spindle-shaped Virus 10

David Andrew Goodman
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the Biology Commons, Genetics Commons, and the Virology Commons

Let us know how access to this document benefits you.

Recommended Citation
Goodman, David Andrew, "Comparative Genetic and Genomic Analysis of the Novel Fusellovirus Sulfolobus Spindle-shaped Virus 10" (2018). Dissertations and Theses. Paper 4496.
https://doi.org/10.15760/etd.6380

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.
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. They provide a malleable and useful genetic tool for probing into the functions of their host, as well as the host responses to infection. Nonetheless, much about these viruses remains to be learned to further understand their morphological, genetic, and life cycle characteristics.

In order to investigate these aspects of these Fuselloviridae, as well as their evolution, this work reports the isolation and characterization of a novel fusellovirus, Sulfolobus Spindle-shaped virus 10 (formerly SSV-L1). Genetic and genomic analyses highlight significant homology 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 anti-CRISPR host evasion. Moreover, we demonstrate the ability of SSV10 to withstand mutation in a fashion consistent with mutagenesis in SSV1. Lastly, analysis of predicted protein structures from SSV10 provide new insights into virus-host interactions. These analyses help to expand our understanding of the viral life cycle while contextualizing the mutagenesis data presented in the following chapters.
Acknowledgements

I would be remiss if I did not begin by thanking my advisor, Dr. Ken Stedman. He allowed me to join his lab at a time when I had little direction, which he quickly and kindly offered me. I will forever be thankful for the exuberant love and enjoyment of his work that was imparted upon me through our interactions; for extending to me so many unique opportunities to contribute to a field that is unlike anything I thought I would ever have a chance to be a part of; for his unending patience in entertaining even my most obtuse questions and ideas, and for being the best mentor a person could ask for. I honestly cannot imagine a more rewarding way to have spent the last six years.

I would also like to give a special thanks to Dr. Eric Iverson who took me under his wing, taught me everything he knew, and made me the competent scientist I am today. You are a great friend, mentor, and brother-in-arms. In addition to molding me into a productive researcher, you taught me how to be an effective mentor as well. Our interactions informed me how to tap into a person’s potential and get the best out of them. My students thank you. I would also like to thank Dr. Geoff Diemer who was the best TA, who invited me to my first Stedman lab meeting, and whose passion inspired me and seeded my love of the world of Archaea and their weird, wacky viruses.

Thank you to all of the people I have had the pleasure of working alongside over the years, especially those of you who went out of your way to help me find success. I have many fond memories, and I will carry them always.
# TABLE OF CONTENTS

| Section | Page |
|---------|------|
| ABSTRACT | i |
| ACKNOWLEDGEMENTS | ii |
| LIST OF TABLES | iv |
| LIST OF FIGURES | v |

## CHAPTER 1: INTRODUCTION

Archaea and the Sulfolobales ................. 1
The Crenarchaeal Virus Family Fuselloviridae .................. 2
A Novel Fusellovirus: Sulfolobus Spindle-shaped Virus 10 .......... 4

## CHAPTER 2: MATERIALS & METHODS

Archaeal and Bacterial Strains .................. 5
Growth Media and Culture Conditions ............ 5
Isolation of Viral DNA ................................ 6
Transposon Mutagenesis of the SSV10 Genome .................. 7
Knockout Mutants of SSV10 .......................... 8
Testing Tn5 Mutants for Infectivity .................... 9
Sequencing and Bioinformatic Analysis .......... 10
Transmission Electron Microscopy ................ 11

## CHAPTER 3: RESULTS & DISCUSSIONS

SSV10 is a Novel Member of the Fuselloviridae .................. 12
SSV10 Mutagenesis .................................. 16
The SSV10 Putative Cas4 Homolog ORF B205 ............... 21
Mutations of SSV10 ORF B205 .......................... 24
Mutations of the Minor Capsid Gene VP3 ................. 28
Synteny of Partially Conserved ORFs ................ 30
MarR-like ORFs in Fusellovirus Genomes ................ 31
PD-(D/E)XK Nucleases Encoded by the Fuselloviridae .......... 35
Relative Conservation of the Putative Tx-Encoded ORFs .......... 36
Sequence Analysis Reveals Conserved Promoter Elements in the Fuselloviridae 37
Sequence Analysis Reveals Non-canonical Promoters and Transcripts .......... 40
Predicted Structures of SSV10 Proteins .................. 46
Structural Predictions Provide Insight into the Life Cycle of the Fuselloviridae 49

## CHAPTER 4: Summary & Future Directions

References .................................. 63
Appendix: List of Primers Used .................. 71
List of Tables

Table 1: Fusellovirus Genomes used in this study ........................................... 6
Table 2: Sulfolobus and E. coli Strains .............................................................. 7
Table 3: SSV10 Mutagenesis ........................................................................ 18
Table 4: SSV10 Putative Promoters ............................................................... 39
Table 5: Phyre2 Protein Structural Predictions for ORFs of SSV10 ............. 53
List of Figures

Figure 1: Transmission Electron Micrographs of SSV10 .......................... 13
Figure 2: Conservation and Mutagenesis of SSV10 ORFs ......................... 14
Figure 3: Halo Morphology of SSV10::Tn5 Mutants ................................. 20
Figure 4: Comparison of Cas4 in SSV10 and S. solfataricus ......................... 23
Figure 5: Halo Assay of SSV10 ΔB205 mutant ...................................... 25
Figure 6a: Halo Assay and PCR Confirmation of SSV10 ΔVP3 mutant .......... 29
Figure 6b: Electron Micrographs of SSV10 and SSV10 ΔVP3 mutant .......... 29
Figure 7: Fusellovirus Genome Comparison .............................................. 34
Figure 8: Fusellovirus Tx Transcript ORF Alignment .............................. 37
Figure 9a: Conservation Among Canonical Fusellovirus Promoters .......... 43
Figure 9b: Conservation Among Non-Canonical Fusellovirus Promoters ...... 45
Chapter 1: Introduction

Archaea and the Sulfolobales

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. Likewise, archaeal viruses — particularly those infecting Crenarchaea — come packaged in strikingly diverse and interesting ways, exhibiting incredibly unique morphologies not known to exist in other domains, including spindle-shaped, bottle-shaped, and droplet-shaped virions. The majority of archaeal viruses also harbor either a linear or circular dsDNA genome, and many archaeal open reading frames (ORFs) encode gene products of unknown function, sharing little or no similarity to known sequences. These Archaea-specific gene compositions may be responsible for archaeal viruses unique methods of entry and egress from their hosts, as well as possibly conferring thermotolerance. Unfortunately, the molecular basis for 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.
The *Sulfolobaceae* is the model thermoacidophilic Crenarchaeal family, found in acidic hot springs worldwide. Members of the family are infected by members of the viral family *Fuselloviridae* (Table 1). Major archaeal biological functions including replication, transcription, and translation have been studied using members of the *Sulfolobaceae* family as models. Furthermore, the mechanisms and capacity of the CRISPR/Cas array employed by Crenarchaeal systems have been extensively studied in recent years, highlighting the complex nature of these adaptive immune systems. The *Sulfolobaceae* utilize a combination of a type I-A acquisition complex along with type I-A and type III-B interference complexes, which have shown differential levels of gene expression and activity in response to an array of substrates and invasive genetic elements. However, very few studies have examined the CRISPR/Cas response in response to direct interaction of *Sulfolobaceae* with their naturally occurring viral predators.

**The Crenarchaeal Virus Family Fuselloviridae**

Currently, *Sulfolobus* Spindle-shaped Virus 1 (SSV1) — isolated from Beppu, Japan — is the best studied member of the *Fuselloviridae*. The complete sequence of the 15.5kb circular, double-stranded DNA genome of SSV1 was determined in 1991 (Table 1). Virus production does not appear to kill host cells but can be induced by UV-irradiation which, within the fuselloviruses, seems to be unique to SSV1. UV induction was utilized in conjunction with
Northern analyses to identify a total of 34 ORFs spanning 11 transcripts. Additionally, the more recent discovery of a novel transcript, T_lys, was reported to be expressed in the absence of UV-irradiation and is thought to play a role in maintaining an uninduced ‘carrier’ state of SSV1. Most recently, SSV1 was shown to be surprisingly tolerant to mutation, with ~50% of ORFs able to tolerate mutation or deletion without completely abrogating activity. This is intriguing as presumably some of these novel genes and their unique characteristics at high temperatures are responsible for both infection mechanics and thermotolerance. Though that study presented genetic conservation of ORFs within the Fuselloviridae, no phylogenetic relationships were analyzed and therefore no comparative genetic data were presented.

Attempting to determine the function of these novel gene products is challenging, however, as they present a near total lack of genomic homology outside of the viral family, with only ORF D335 — a viral integrase of the tyrosine recombinase family — having significant sequence similarity to previously characterized proteins. Nevertheless, structural studies and proteomic analyses have illuminated various functional aspects of the life cycle of viruses of Archaea.

Though the replication process of the Sulfolobales has been well described, these observations have not provided much insight into the replication origins and processes associated with replication of viral genomes in vivo. Nevertheless, despite its lack of similarity to the ori sequences of its hosts,
an intergenic region in the SSV1 genome has been implicated as a replication origin and shown to allow replication of a shuttle vector in *Sulfolobus*\textsuperscript{10}. Unfortunately, this region is not obviously conserved across the *Fuselloviridae* and the specific mechanisms and interactions responsible for replication initiation in this region remain unknown.

**A Novel Fusellovirus: Sulfolobus Spindle-shaped Virus 10**

This work furthers understanding of archaeal viruses, presenting a comparative genetic and genomic analysis of the novel Sulfolobus Spindle-shaped Virus 10 (formerly SSVL-1) with other members of the *Fuselloviridae*, indicating that there are some shared genes critical for virus function. Originally isolated from a hot spring in Lassen Volcanic National Park in northeastern California, the novel Sulfolobus Spindle-shaped Virus 10 (SSV10) encapsidates a 16.3kbp circular, double-stranded DNA genome sequence. SSV10 encodes 40 ORFs, of which 5 share no homology to any other ORFs within the viral family. Moreover, SSV10 has a variable host range, infecting 6 of 13 different species of *Sulfolobus*, 3 more than SSV1\textsuperscript{11}. SSV10 also encodes a putatively functional Cas4-like ORF, which may dictate this host range and play a role in anti-CRISPR host evasion. An in-depth analysis of genetic relationships within the *Fuselloviridae* is also reported. Lastly, SSV10 is shown to withstand transposon and deletion mutagenesis in a fashion consistent with mutagenesis in SSV1\textsuperscript{41}. 
Chapter 2: Methods

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\(^{11}\), and is susceptible to infection by multiple wild-type SSVs as well as an array of SSV1 deletion and insertion mutants\(^{41}\). All *Sulfolobus* strains, both infected and uninfected, were isolated as in Zillig et al. (1993)\(^{109}\). Briefly, single colonies were isolated from enrichment cultures and screened for virus production by spot on lawn assays of lawns 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+ *E. coli* (Epicentre) was used as host for transformations of transposon-containing constructs.

Growth Media and Culturing Conditions

All *Sulfolobus* strains, both infected and uninfected, were cultured aerobically in Yeast-sucrose (YS) media or on Gelrite plates at 75-80°C as previously described\(^{40}\). Media pH was adjusted to between 3.0-3.2 using 50% H\(_2\)SO\(_4\) 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\(^{33}\). Table 2 contains a complete list of the strains used in this study and their genotypes.
Table 1: Fusellovirus genomes used in this study

| Virus Genome (Location)* | Size (bp) | Annotated ORFs | Accession Number | Reference |
|--------------------------|-----------|----------------|------------------|-----------|
| ASV1 (ISL)               | 24,186    | 38             | NC_013585        | 82        |
| SSV1 (JPN)               | 15,465    | 35             | NC_001338        | 66        |
| SSV2 (ISL)               | 14,795    | 35             | NC_005265        | 100       |
| SSV3 (ISL)               | 15,230    | 32             | KY579375         | 97        |
| SSV4 (ISL)               | 15,135    | 34             | NC_009986        | 70        |
| SSV5 (ISL)               | 15,330    | 34             | NC_011217        | 82        |
| SSV6 (ISL)               | 15,684    | 33             | NC_013587        | 82        |
| SSV7 (ISL)               | 17,602    | 33             | NC_013588        | 82        |
| SSV8 (USA)               | 16,473    | 37             | NC_005360        | 103       |
| SSV9 (RUS)               | 17,385    | 31             | NC_005361        | 103       |
| SSV10 (USA)              | 16,271    | 40             | KY563228         | This Work |

*(Location): ISL, Iceland; JPN, Japan; RUS, Russia.

Isolation of Viral DNA

SSV10 episomal DNA was isolated from Sulfolobus strain S355 first by alkaline lysis, then extracted with phenol:chloroform:isoamyl alcohol (25:24:1) followed by ethanol precipitation. DNA extraction was confirmed via agarose gel electrophoresis after restriction endonuclease digestion with EcoRI (New England Biolabs) before being further purified using a GeneJet Plasmid Miniprep Kit (Thermo-Fisher).
Table 2: *Sulfolobus* and *E. coli* Strains

| Strain            | Description/Genotype                                                                 | Reference |
|-------------------|---------------------------------------------------------------------------------------|-----------|
| *E. coli* EC100D pir<sup>+</sup> | F<sup>−</sup> mcrA Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ<sup>−</sup> rpsL (Str<sup>3</sup>) nupG pir+(DHFR) | Epicentre, Inc |
| *S. solfataricus* S441 | *S. solfataricus* isolate; SSV host                                                   | 11        |
| *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 53,110 |
| *S. solfataricus* P2 | *S. solfataricus* isolate with complete genome sequence                               | DSM 1617 95,110 |
| *S. solfataricus* GΘ | *S. solfataricus* MT4 derivative lacS                                                 | 10        |

**Transposon Mutagenesis of the SSV10 Genome**

The EZ-Tn5™ <R6Kori/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. (2017), the manufacturer’s recommended equimolar ratio of target DNA to EZ-Tn5 transposon was replaced with a 30:1 molar ratio<sup>+</sup>. SSV10::Tn5 reactions were transformed into Transformax EC100D pir+ *E. coli* and grown on LB agar plates with 50 µg/mL kanamycin. Plasmid DNA purified from single colonies was
isolated and screened for mutagenized SSV10 DNA via EcoRI or EcoRV endonuclease digestion and agarose gel electrophoresis.

**Knockout Mutants of SSV10**

Deletions of SSV10 ORFs were made using long-inverse PCR (LIPCR) as previously described\textsuperscript{18,40,41}. The $T_m$ calculator software (http://tmcalculator.neb.com/#/) was used to estimate the annealing temperatures for each primer pair (see Supplemental Table 1 for primer sequences). 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 150-250 ng/µl, was initially diluted 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 0.005U/20µl as described previously\textsuperscript{41}. 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. LIPCR generated plasmid DNA was purified as described above.
**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)$^{92}$. 50 mL of cells were pelleted (15 min. @ 3000 g) and washed in decreasing volumes of 20 mM sucrose to a final volume of four hundred microliters with a final concentration of cells at $\sim 10^{10}$ cells/ml. One hundred microliters of washed cells were added to a chilled 0.1-cm-gap-length cuvette (VWR), and 2μl of SSV10 or SSV10::Tn5 DNA (~300-1000 ng total DNA) was added to the cells.

Transformations were performed via electroporation (Gene Pulser II; Bio-Rad) at 1.5 kV, 400Ω, and 25μF. 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 hour at 75°C. Following recovery, cells were transferred to 50 mL of preheated YS in long-neck Erlenmeyer flasks and grown with shaking from 74-78°C.

Spot-on-lawn (halo) assays were performed as described previously$^{40}$. Briefly, five microliters of transformed cultures were spotted on a lawn of uninfected *Sulfolobus* at mid-logarithmic growth (500μl/plate at OD$_{600} = 0.2 - 0.5$) in a 0.2% YS Gelrite soft-layer poured over a 1% 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 for transformation. Negative controls were uninfected
Sulfolobus cultures. To confirm the identity of the viral DNA in cultures that inhibited host growth by halo assay, viral DNA was purified from transformed Sulfolobus cells and amplified via PCR using specific primers (Supplemental Table 1) flanking the mutated region of the viral DNA. Control PCRs were done using the DNA used for transformation and wild-type SSV10 DNA.

**Sequencing and Bioinformatic Analysis**

Sanger sequencing was performed using transposon-specific primers on 172 SSV10 plasmids containing the EZ-Tn5™ <R6Kyor/KAN-2>Tnp transposon. Ambiguous sequences were corrected by re-sequence and gaps were closed using custom primers (Supplemental Table 1). Sequencing was performed either by Eurofins© Genomics or the DNA Services Core at the Oregon Health & Science University. Geneious™ V. 8.1 (Biomatters) was used to identify ORFs in SSV10, and for generating nucleotide and protein alignments using the Geneious™ Needleman-Wunsch alignment tool with either a BLOSUM62 or BLOSUM80 substitution matrix depending on query length. Similar sequences to SSV10 ORFs were identified in the NCBI database using the Basic Local Alignment Search Tool (BLASTP)². Searching for CRISPR spacers in Sulfolobus genomes was done using the BLAT search engine provided by the UCSC Archaeal Genome Browser¹²,⁴³. Specific start codons for each ORF were determined using a combination of the best fit per BLASTP e-value, likely usage rate⁶⁹,¹⁰² of any particular start codon, and the presence of canonical Sulfolobus
ribosome-binding sites\textsuperscript{102}. Putative transcripts were identified by the presence of predicted upstream promoter elements, identified by their respective TFB-recognition elements (BRE) with adjoining TATA box\textsuperscript{26,28,45,85}. Structural models of putative SSV10 proteins were generated using ‘intensive’ modelling by the Protein Homology/analogY Recognition Engine V. 2.0 (Phyre2) web portal for protein modeling, prediction and analysis using otherwise default parameters\textsuperscript{42}.

\textbf{Transmission Electron Microscopy}

For transmission electron microscopy, samples were prepared on 400-mesh carbon-Formvar-coated copper grids (Ted Pella, Redding, CA, USA). Grids were placed, carbon-Formvar down, on a 5 μL droplet of culture supernatant for 2 min. Culture supernatants were generated by centrifugation at 6000 g for 5 min. Samples were removed from the grid by wicking. Grids were then stained for 15–60 s on 5 μL of 2% 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 8500 to 34,000 magnification on an FEI Tecnai F20 transmission electron microscope (TEM) (FEI Inc. Hillsboro, OR, USA).
Chapter 3: Results & Discussion

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\(^{11,17}\). Episomal DNA, originally SSV-L1, now SSV10\(^{41}\), purified from S355 was transformed into the laboratory *Sulfolobus* host S441 via electroporation\(^{92}\), generating *Sulfolobus* strain S592 (Table 2). Strain S592 produced infectious virus via halo assay, virions by transmission electron microscopy (Figure 1), and subsequent recovery of SSV10 DNA from transformed cultures\(^{40}\), confirming that this represented the virus genome. The GC content of SSV10 is 37.1 %, consistent with the 38.5% median of Fusellovirus GC content. The SSV10 genome contains 40 ORFs and 10 predicted transcripts (Figure 2). ORFs in the SSV10 genome (Accession number KY563228) were annotated 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 (Figure 2). TFB-recognition elements (BRE) and corresponding TATA boxes found in SSV1\(^{26,28,85}\) were used to annotate 10 putative transcripts in SSV10.
Figure 1: Negative stain transmission electron micrographs of Sulfolobus Spindle-shaped virus 10. Viral particles are similar in size and morphology to most other members of the Fuselloviridae.

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 universally conserved ‘core’ genes and variable genes, with most of the variable genes in the T5 and T6 ‘early’ transcripts\textsuperscript{26,85}, similar to other SSVs (Figure 2). SSV10 has an average nucleotide sequence identity of ~41\% within the family and is most similar to Yellowstone National Park’s SSV8 (formerly SSVRH), the only other published SSV isolated from North America\textsuperscript{103}, at 54.7\%.
Figure 2: Conservation and Mutagenesis of SSV10 ORFs

Overall conservation (BLASTP E-value ≤ 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 SSV10 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. Virus capsid genes and the integrase gene are labeled in parentheses. Putative transcripts are labeled with a capital T. Thin arrows represent the location of Tn5 insertion mutants characterized as functional (green) and nonfunctional (red). ORFs are labeled as in Genbank Accession KY563228.1.
Conservation of ORFs ranges from 12 completely conserved ‘core’ ORFs to 5 unique to SSV10 (Figure 2). About 68% 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 5 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’-UTRs, translational start sites may or may not correspond to the first start (AUG, GUG, 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.

As noted previously by Held and Whitaker (2009), whole genome and individual ORF phylogenies of fuselloviruses do not have consistent branching patterns. Gene synteny and conservation of single ORFs indicates that ASV1, SSV7 and SSV9 are divergent from other SSVs. Sequence analysis of the
ORF D154 MarR-like putative transcriptional regulator also indicates that the ASV1, SSV7 and SSV9 fuselloviruses differ from other SSVs.

**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. 2017\(^41\) (Figure 3; Table 3). SSV10 is genetically tractable, tolerating both insertion and deletion mutations in a manner similar to SSV1\(^18,40,41,99\). A mutant with an insertion in the SSV10 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% 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% efficient\(^41\).

Unique insertion mutations were obtained in 35 out of 40 SSV10 ORFs, including all ORFs in the putative T5 and T5’ transcripts. 16 different mutants in 14 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 3). Two deletion mutants, DAG593\(_\Delta_B\)205 (DAG821) and DAG593\(_\Delta_VP3\) (DAG825), were also
generated using Long Inverse PCR (LIPCR)\textsuperscript{18}. All functional SSV10::Tn5 insertion and deletion mutants were confirmed via DNA extraction from transformed \textit{Sulfolobus} and subsequent PCR amplification of the mutagenized region (see Supplemental Table 1 for primers). Consistent with mutagenesis in SSV1, SSV10 ORFs encoded on the T5 and T5’ transcripts generally tolerate transposon insertion without abrogating function. In total, 8 of 16 transposon mutants in the T5 and T5’ transcripts were functional. Most of these insertions were in ORFs containing putative DNA binding domains characteristic of transcription regulatory proteins. Furthermore, all 8 of these functional mutants occurred in the 3’-end of the T5 transcript, or in the two ORFs in the T5’ transcript (Figure 2). Mutants in both ORFs of the novel T5’ transcript were functional, which is not surprising considering that they are relatively poorly conserved. SSV10 ORF F73 in the T5’ transcript is a predicted copG-like RHH 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\textsuperscript{41}. Starkly different from SSV1, homologues of these ORFs in SSV2 were the last to be upregulated\textsuperscript{86} late in the replication cycle, 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 copG-like RHH products in the T5 and T6 transcripts, respectively, and are expressed late (~4.5 - 6 h.p.i.) in SSV2\textsuperscript{86}. 
Table 3: SSV10 Mutagenesis

| Plasmid | Position | ORF   | Transcript | Positive | 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) | Tint | 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   | Tori       | 2        | 1        | Y          |
| DAG702  | 4468     | A49   | Tori       | 0        | 3        | N          |
| DAG719  | 3203     | INT (D335) | Tint | 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
The four SSV10 ORFs starting from the 5’-end of the T5 transcript — E62, D154, F111, and D152 — did not tolerate transposon insertion, which lends credence to predictions about their respective roles and importance in the viral life cycle (Table 5). SSV10 ORF D154 is predicted to be a MarR-like transcriptional repressor (Figure 2; 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, or are potentially host dependent.

Insertions in the SSV10 putative 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\textsuperscript{40}. Halos generated by SSV10 ORF F199 insertion mutants (DAG593, DAG698) were similar to those generated by wild-type SSV10 (Figure 3), and viral titers quantified by number of plaque forming units (PFU) via plaque assay (~1.3x10\textsuperscript{4} to 4.5x10\textsuperscript{5} PFU/ml) are equivalent. Two mutants in SSV10 ORF E52 were characterized; one was functional while the other was not. The insertion in the functional SSV10 ORF E52 mutant is located just one residue downstream of the annotated ATG start, yet there are two more in-frame ATG start codons 3 and 7 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 is likely critical for virus function.

**Figure 3: Halo Morphology of SSV10::Tn5 Mutants**

*Figure 3: Two examples of halo formation caused by the spotting of cultures transformed with SSV10::Tn5 insertion mutants. Halos generated by DAG593 (left) and DAG698 (right) are comparable to those formed by the wild-type (+) control.*

Insertions in the SSV10 integrase gene, including one nearly identical to a functional mutant of SSV1\textsuperscript{41}, 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. 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.

The SSV10 Putative Cas4 Homolog ORF B205

Sequence similarity searches and protein alignments done using BLASTP (NCBI) on the SSV10 ORF B205 predict a CRISPR-associated Cas4-like protein product (E-value = 4e-10) that is homologous to ORFs present in 8 out of 11 members of the Fuselloviridae but is not present in SSV1. Structural predictions created using Phyre2 show high confidence (>90%) matches with various known archaeal and bacterial Cas4 and RecB protein structures, most notably with the Cas4 SSO0001 encoded by Sulfolobus solfataricus (Table 5). These proteins form a decameric toroidal quaternary structure thought to be part of the CRISPR/Cas spacer acquisition complex along with Cas1 and Cas2. Moreover, SSV10 ORF B205 and all other SSV-encoded Cas4-like ORFs contain the 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’→3’ exonuclease activity (Figure 4). However, it is currently unknown whether these virally
encoded Cas4 genes are capable of forming higher order structures leading to unwind ing and exonuclease activities similar to their host-encoded counterparts.

Interestingly, Cas4 homologues 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 been repurposed as the nucleocapsid protein TP1. 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.

* Sulfolobus* genomes encode CRISPR arrays combining the well conserved type I-A acquisition complex 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*. Cas4 has been shown to form a complex with Cas1 and Cas2 or fused to Cas1 directly. 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, which cannot be ruled out as it relates to SSV10.
Figure 4: Comparison of Cas4 in SSV10 and \textit{S. solfataricus}

Figure 4a: The predicted structure of the Cas4-like gene product from SSV10 (left) compared to a the crystal structure of SSO0001 monomer from \textit{S. solfataricus} (right). The quad-cysteine [4Fe-4S] binding site residues are in yellow, while the active site residues are colored by residue in the center of the structure. SSO0001 is rotated ~45° on the vertical access to better show the active site.

Figure 4b: The putative Cas4-like genes of the Fuselloviridae aligned to SSO0001 of \textit{S. solfataricus}. Identical amino acid residues are highlighted. Active site residues and [4Fe-4S] cysteines are marked with gold stars.
Mutations of SSV10 ORF B205

Due its similarity with the host-encoded Cas4, mutants in SSV10 ORF B205 were tested (Table 3). Two Tn5 insertion mutants at or near codons for active site residues within B205 (Figure 4b) 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 predicted N-terminal portion of the conserved downstream SSV10 ORF C127, homologues of which were shown to be necessary for virus function in SSV141. A deletion of SSV10 ORF B205 (DAG821, Table 3) was generated using long-inverse PCR (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 when spotted on lawns of uninfected Sulfolobus after 11 separate transformations and were thus characterized as non-functional (Table 3). Conversely, cultures transformed with the deletion mutant DAG821 (ΔB205) generated halos in 3 out of 8 trials. Sulfolobus infected with DAG821 exhibit a phenotype consistent with an active host CRISPR/Cas response to infection58. Spot-on-lawn assays performed at 72 and 96 hours post-transformation generated halos, indicating production of infectious particles in the transformed culture. Halos produced 96 hours post-transformation were significantly diminished compared to halos at 72
hours post-transformation. At 120 hours post-transformation no observable halos were produced (Figure 5). Nonetheless, PCR amplification of DNA purified from transformed cells showed that SSV10 viral DNA lacking ORF B205 was still present in the culture after one 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.

**Figure 5: Halo Assay of SSV10 ΔB205 mutant**

*Figure 5: 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 hours post transformation (h.p.t). (Bottom) Halos of SSV10 lacking ORF B205 are consistent with the (+) control at 72 h.p.t. However, infectivity is greatly reduced at 96 h.p.t, and completely lost by 120 h.p.t.*
If SSV10 ORF B205 and its homologues are indeed active Cas4 proteins, they may represent another class of archaeal virally-encoded anti-CRISPR/Cas, similar to mechanisms which aid bacterial viruses in evading the host CRISPR/Cas system\textsuperscript{7,8,16,39,67,68,94}. A deletion of SSV10 ORF B205 (DAG821), generates a functional mutant. 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 DAG821. 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.

By contrast, insertion mutations in SSV10 ORF B205 lead to loss of function. In a few cases, the loss of function in insertion mutants has been attributed to polar effects\textsuperscript{41}, whereby the interruption of a specific gene product via transposon perturbs the downstream transcription or translation of other gene products. SSV10 ORF C127 is a highly conserved gene downstream of ORF B205, and the homologue in SSV1, ORF B129\textsuperscript{63}, has previously been shown not to tolerate deletion or insertion mutations\textsuperscript{40,41}, and is thus required for virus infectivity. It may be that the loss-of-function phenotype exhibited by insertion mutants of ORF B205 are actually due to polar effects disrupting ORF
C127 which would explain the disparity between phenotypes of insertion and deletion mutants of SSV10 ORF B205.

The laboratory host *S. solfataricus* S441 used in this study has recently been sequenced and CRISPR arrays identified (Unpublished data). Despite both S441 and SSV10 being isolated from Lassen Volcanic National Park, no CRISPR spacers with matches to SSV10 have as of yet been identified in any putative CRISPR locus of *S. solfataricus* S441. This may be an indication that the putative anti-CRISPR capability of SSV10 presented in this work is active and prevents acquisition of SSV10 spacers by S441. *S. solfataricus* S441 is highly similar to *S. solfataricus* P2 (GenBank ID: AE006641), which carries two significant matches to SSV10 in CRISPR array #11 seemingly absent from S441. Despite the presence of these spacers in *S. solfataricus* P2, SSV10 is still able to sustain an infection after transformation. 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 CRISPR spacers\textsuperscript{22,30,31,82}. Characterizing the infection cycle and host range of SSV10 ΔB205 in comparison with both wild-type SSV10 as well as with an SSV lacking a Cas4 homologue altogether (such as SSV1) will further illuminate the nature of the genetic arms race between the *Fuselloviridae* and its hosts and may ultimately lead to the unraveling of a novel viral anti-CRISPR defense mechanism in the archaea.
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 relative to Tn5 insertions in SSV141. Both mutants were found to be nonfunctional (Figure 2; 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 mutants lacking VP3 exhibit reduced halo formation on lawns of S.solfataricus (Figure 6a) and elongated virions on TEM analysis (Figure 6b), similar to deletions of VP3 in SSV141. 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 VP341. However, Tn5 insertions in SSV10 VP3 are maintained — likely due to a lack of similar direct repeats in SSV10 — and lead to a complete loss of infectivity (Figure 2). Why an insertion in VP3 is not tolerated while a deletion is tolerated is not clear. VP3 falls at the end of its respective transcript where, unlike insertion mutants in ORF B205, there are not likely to be polar effects on other ORFs or transcripts required for functional virus infection. It may be that a partial, defective VP3 protein is formed that disrupts assembly of mature virions, which could explain evolutionary pressure to delete insertions in SSV1 VP341.
Figure 6a: Halo Assay and PCR Confirmation of SSV10 ΔVP3 mutant

Figure 6a: (Left) Spot-on-lawn assay of SSV10 VP3 deletion mutants. Three separate transformations of DAG825 (DAG593_ΔVP3) are spotted, with transformations #1 and #3 retarding growth of the uninfected S441 lawn. S441 infected with SSV10::DAG593 is the (+) control, uninfected S441 are the (-) control. (Right) PCR confirmation of infected cultures. Lane 1, Axygen Biosciences 1kb Ladder; lane 2 and 3, DAG825 DNA purified from S441; lane 4, DAG825 purified from E. coli; lane 5, no DNA negative control; lane 6, DAG593 purified from S441

Figure 6b: Electron Micrographs of SSV10 and SSV10 ΔVP3 mutant

Figure 6b: (Left) Transmission electron micrograph of negatively-stained SSV10 wild-type virus particles. Red arrows indicate Fuseliovirus-like 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 consistent with ΔVP3 mutants in SSV1. Micrographs are typical for SSVs.
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 (Figure 2), are accompanied by a promoter upstream from the F73 start codon. A homologous coding region along with highly conserved promoter elements (Figure 9b), is present in 7 of 11 Fuselloviridae. Microarray analysis of the SSV2 infection cycle supports this annotation, showing upregulation of these homologues in tandem ~8-9 hours post infection [86]. 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 (Figure 7).

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 intra-genome recombination [82] 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 [11] and appears to induce a dormant or apoptotic state in infected host cells [5] rather than establishing a stable carrier state similar to other SSVs. ASV1 and SSV7 have not been studied in depth, but
ASV1 infects *Acidianus* not *Sulfolobus*, a further indication of potential host range changes due to rearrangement.

**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\(^1\). In *Escherichia coli*, this is accomplished by binding to the marO promoter, negatively regulating the marRAB operon\(^1\). Alternatively, the crystal structure and DNA binding mechanisms of a *S. solfataricus* MarR homologue BldR\(^23\) 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*, BldR\(^24\), 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, a MarR-like transcriptional regulator encoded in the putative 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\(^46,48\). The MarR-like ORFs encoded by ASV1,
SSV7, and SSV9 are also truncated, missing between 50-70 amino acids at the N-terminal end relative to other SSVs but maintain a putative helix-turn-helix or winged-helix domain, as well as a conserved [DXIIYXLTEKG] C-terminal motif also present in BldR and BldR2 from S. solfataricus. These truncated ORFs are present at a genetically distinct location between the T5\textsuperscript{+} and T6 transcripts, similar to the position of the non-conserved T\textsubscript{lys} transcript in SSV1. 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\textsubscript{lys} promoter of SSV1.

SSV2 encodes a seemingly bipartite version of the full-length MarR-like ORF made up of ORF88a and ORF 106. Transcript analysis of SSV2 shows that ORFs106 and 88a, separated by ~1200 base pairs are expressed in tandem between 2 and 3 hours post infection\textsuperscript{86}. Structural predictions indicate many high confidence matches of the MarR-like winged helix-turn-helix DNA-binding domain to SSV2 ORF106, homologous to the C-terminal end of the full-length MarR-like ORFs found in other Fuselloviruses. On the other hand, SSV2 ORF88a shares only a few conserved residues associated with the N-terminal end of other SSV-encoded MarR-like ORFs. However, structural predictions reveal high confidence matches to two Crenarchaeal viruses of the Ligamenvirales\textsuperscript{75}: SIFV ORF14\textsuperscript{32} and SIRV3 AcrID1\textsuperscript{35}. Interestingly, both of these homologues form homodimers, and SIFV ORF14 in particular seemingly lacks a
catalytic domain\textsuperscript{32}. This homology and lack of catalytic function lends credence to the idea that SSV2 ORF88a, and its homologue in SSV6 ORF B87, may be dimerizing and binding ORF106 in order to form a heterotetramer which acts as a functional homologue to other MarR-like transcriptional regulators. Moreover, these ORFs in both SSV2 and SSV6 contain two short direct repeats which may be indicative of a split or would allow for them to become concatenated should a junction form.

Curiously, a homologue of SSV2 ORF88a is found in SSV3 which itself already encodes a full length MarR-like gene. The homologue AcrID1 encoded by SIRV3 has been implicated in anti-CRISPR activity, and the underlying structural mechanisms as well as their widespread homology across Archaeal virus families has recently been reported\textsuperscript{35}. If this is indeed the case for SSV3, it may provide an alternative to the function postulated above, and why it would be maintained in SSV3 where a full length MarR-like gene is already present. Overall, SSV10 ORF D154 and its diverse homologues are further evidence that Fusellovirus genes once thought to be less conserved may provide another way for the ancestral Fusellovirus lineage to be traced.
Figure 7: 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. Open reading frames (ORFs) in each genome are shown as outlined arrows. ORFs in transcripts are color-coded, with promoters 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 grey are those outside of the labeled transcripts.
**PD-(D/E)XK Nucleases Encoded by the Fuselloviridae**

SSV10 ORF F199 encodes a predicted novel viral nuclease (E-value = 1.45e-12). Primary amino acid analysis of SSV10 ORF F199, in comparison with the other predicted viral nucleases encoded by the *Fuselloviridae*, reveals a conserved PD-(D/E)XK catalytic motif in a putative viral nuclease that is otherwise starkly different those found in the rest of the *Fuselloviridae*.

Interestingly, SSV10 ORF F199 is homologous to a host-encoded nuclease found in both *S. solfataricus* strain P2 as well as strain S441, sharing nearly 50% identity over 96% of the primary sequence. The amino acid sequence shares little similarity, about 25% identity, to that of the 7 other members of the *Fuselloviridae* nucleases. Despite this poor similarity with the family overall, ORF F199 shares 30.5% identity with its putative homologue in SSV5. By contrast, while the other 6 nucleases in the family share a highly conserved consensus sequence at nearly 67% identity, they are much more diverged from their closest structural homologues from *S. solfataricus* (*Sso Hjc*, PDB ID 1HH1)\(^2\). It may be that, while the catalytic motif is conserved throughout the family, the other portions of the protein may play a role in site-specific recognition that is dependent on the target. Another possibility is that SSV10 ORF F199 was acquired from the host genome through a recombination event, either via integration and subsequent excision of prophage in a region near the host-encoded homologue or through the catalytic mechanism of the host-encoded gene product.
**Relative Conservation of the Putative Tx-Encoded ORFs**

The transcript Tx\textsuperscript{26} (Figure 2, Figure 8, Figure 9b) that was not originally annotated in SSV1\textsuperscript{83,85} appears to be conserved in all but one SSV. Moreover, the ORFs themselves both encode conserved motifs shared by other C124 homologues. These homologues seem to come in three types based off of conserved motifs along the length of the primary protein sequence (Figure 8).

The first group are all over 120 amino acids, and while they share some conserved residues with C124 of SSV1, ASV1 and SSV2 are clearly more distantly derived and lack any significant homology with the other fuselloviruses. C124 is not itself highly similar to the homologues, only maintaining a C-terminal [PPWYANLWPEVLTIGI] motif conserved in SSVs 3-5 and SSV9. These four SSVs make up the second type, with SSVs 3-5 being completely identical and SSV9 matching 86\% of residues. These ‘mid-length’ homologues are typified by two unique N-terminal motifs not found elsewhere. The last group consists of the four shortest ORFs, formed from SSV7, the duplicate ORFs of SSV8, and SSV10. These ‘short’ homologues maintain a C-terminal motif congruent with the other homologues, however they share a unique N-terminal motif which differs from that of the ‘mid-length’ homologues.
Figure 8: Fusellovirus Tx Transcript ORF Alignment

Figure 8: An alignment of ORFs encoded on the Tx Transcript of each Fusellovirus used in this study. The shortest ORFs are boxed in red, the longest are boxed in blue, and the mid-length ORFs are not boxed. SSV3 ORF33 is representative of SSV4 and SSV5 as all three are identical. Consensus residues are highlighted.

Sequence Analysis Reveals Conserved Promoter Elements in the Fuselloviridae

For all SSVs, there are ‘core’ transcripts corresponding to the SSV1 and SSV10 (Table 4) transcripts T1/2, T3, T4/7/8, and T9 (Figure 9a). 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 7th position. Conversely, the TFB-recognition elements of the T1/2 transcripts seem distinct from that of the other ‘core’ transcripts. The BREs of all 4 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 AGGCC, while 8 out of 11 of the T1/2 BREs share a consensus sequence of SRGGGG (Figure 9a).

Promoters for the T5 and T6 ‘early’ transcripts\textsuperscript{96,83} 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 (Figure 9a). Variations between promoters in different SSVs seem to be between putative TFB-recognition elements (BREs). Previous studies\textsuperscript{69,81} 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. Three of the \textit{Fuselloviridae} — ASV1, SSV7, and SSV9 — lack a polycistronic T5-like negative strand transcript altogether. Instead, these genomes have a putative positive strand transcript in the corresponding part of the genome, referred to herein as T5\textsuperscript{+}, with large open reading frames including a rad3-like helicase not found in other members of the \textit{Fuselloviridae} (Figure 7).
## Table 4: SSV10 Putative Promoters

| Transcript | Promoter Sequence* | SSV1 Match (BRE/TATA) | Features |
|------------|--------------------|----------------------|----------|
| T1/2       | CTGAATTCAGAACTAGGGGGGT TTAAAAGCTTAGTGAAGCC CTTATTAAACAGGAGT | Y/Y      | One mismatch in BRE |
| T3         | ATTCGTAATGCACTCTTTTTTATTAG GCCTTTTAAAGTTACACTTTCC TTTTTCGCTTACA4ATG | N/N      | Matches T9 promoter |
| T4/7/8     | TTCCTCGTAAGACGAAAAATAGAT TAAGCCCTTTTAAAGTCACATA ATTTTATCGCTTACAATG | Y/Y      | One mismatch to T9 BRE |
| T5         | AGAAAGAGAGATAGATGAACG AAAGATTATATAGTCAGATAGA TAGATAGATAAAATG | Y/N      | One mismatch in BRE; No Repeats |
| T6         | TCttgatagattgatagattgaGAAAGAATTATATA TGCACTTTTTG ATATGTTTAATGAAATG | Y/N      | One mismatch in repeats, BRE |
| T9         | GTATAAAATCTACAAAGACTGAG TAGGCCCTTTTAAAGTCATTTTT CTTTTTCTCATTACAATG | Y/Y      | Most highly conserved |
| Tx         | TACCACATATGCACTCTAAAGGC AAAATTTTATATCCTTTTCTTTTTT ATATGGTTTAATCAATG | N/N      | - |
| Tint       | GATAAGATAATTATATATCACAATTTTT AAATGCTTACGTGATAATATTAA ATGGGCTGAAGGTATG | -        | Novel transcript |
| Tori       | AAAACTTTATACCACATACCTTC AGGCCAATTTATATTACGTGATA AGCATTAAAAGGGATG | -        | Novel transcript |
| T5’        | TTCAGTCCCTCGTTTTTTATGTC CTTTTTTTTGCTATAAAATCTTGT GATAGAGAAGGATG | -        | Novel transcript |

*TFB-Recognition Element in bold; TATA-box underlined; Start/RBS in italics; T6 direct repeats in lowercase

The promoter elements associated with the Tx transcript are highly conserved in every fusellovirus with the exception of SSV6, sharing a consensus
sequence of \texttt{AAAAATTTTTAAAC} (Figure 9b). 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 \textit{Fuselloviridae} and probably regulated similarly. 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.

\textit{Sequence Analysis of SSV10 Reveals Non-Canonical Promoters and Transcripts}

Analysis of potential transcription start sites and the divergent transcripts in SSV10 (Figures 2, Figure 9b) indicates that all SSVs except SSV1\textsuperscript{26} 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 exhibit a eukaryal-like TATA box with a consensus sequence of \texttt{TWTTTAAC} (Figure 9b). 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\textsuperscript{86}.

SSV10 also appears to contain alternative transcripts relative to SSV1, including a unique positive strand transcript $T_{ori}$, and a short transcript upstream of the T5 transcript, herein referred to as T5', which encodes two ORFs — F73
and D60 (Figure 2). The T5’ transcript is found in six other SSVs (Figure 7), the products of which were shown to be the last to be upregulated in SSV2 9 hours post infection of *S. solfataricus* P2\(^86\). 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*\(^4\) (Figure 9b). The T\(_{ori}\) transcript in SSV10 does not have a conserved BRE or TATA box, but the region encoding ORF A49 does contain sequences highly similar to Origin Recognition Boxes (ORB elements) found in the *Sulfolobales*, which have been shown to be well conserved across the Archaeal kingdom and are specifically recognized by the DNA replication protein Orc1/Cdc6\(^88\). Both functional and nonfunctional Tn5 mutants interrupting ORF A49 were generated (Figure 2), and the location of the nonfunctional mutant may provide evidence of a specific origin site. Furthermore, while the genomic location differs, SSV1 does contain similar ORB elements within its predicted origin of replication\(^10,26\). Although no SSV origins have been functionally characterized to date, the presence of these ORB elements provides the first connection to the specific replicative mechanisms of their host cells.

The T\(_{lys}\) transcript that has been reported to be important for maintenance of the carrier state of SSV1\(^28,29\) 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 — homologous to ORF
D57 of SSV8 — which shares no similarity in sequence, promoter elements, or structure with the predicted copG-like transcriptional regulator F55 of SSV1\textsuperscript{29}. Truncated, but well-conserved homologues of SSV10 ORF E56 are also present but unannotated in 5 other SSVs. In ASV1, SSV7, and SSV9, this coding region instead corresponds to a MarR-like HTH transcriptional regulator similar to those encoded in the T5 transcript (Figure 7) of all other SSVs. Unfortunately, no transcript mapping has been done other than for SSV1\textsuperscript{26,85} and SSV2\textsuperscript{86} so all other promoter annotations are speculative.
Figure 9a: Conservation Among Canonical Fusellovirus Promoters

| Promoter Type          | BRE | TATA Box |
|------------------------|-----|----------|
| **T1/T2 Consensus Promoter** |     |          |
| SSV10 T1 Promoter      | GAGGGGTTTTTTAAAAAA | |
| SSV1 T1 Promoter       | GAGGGGTTTTTTAAAAAA | |
| SSV2 T1 Promoter       | GAGGGGTTTTTTAAAAAA | |
| SSV7 T1 Promoter       | GAGGGGTTTTTTAAAAAA | |
| SSV8 T1 Promoter       | GAGGGGTTTTTTAAAAAA | |
| SSV9 T1 Promoter       | GAGGGGTTTTTTAAAAAA | |
| SSV3 T1 Promoter       | CGGGGGTTTTTTAAAAAA | |
| SSV5 T1 Promoter       | CGGGGGTTTTTTAAAAAA | |
| SSV4 T1 Promoter       | ACGAGGTTTTTTAAAAAA | |
| SSV6 T1 Promoter       | TGGGCTTTTTTTAAAAAA | |
| ASV1 T1 Promoter       | ACACACTTTTTTTAAAAAA | |
| **T3 Consensus Promoter** |     |          |
| SSV10 T3 Promoter      | AAGGCCCTTTTTAAAAAA | |
| SSV7 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| SSV9 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| SSV3 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| SSV4 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| SSV5 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| SSV8 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| SSV2 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| ASV1 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| SSV1 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| SSV6 T3 Promoter       | AAGGCCCTTTTTAAAAAA | |
| **T4/7/8 Consensus Promoter** |     |          |
| SSV10 T8 Promoter      | AAGGCCCTTTTTAAAAAA | |
| SSV3 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
| SSV4 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
| SSV5 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
| SSV8 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
| SSV2 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
| SSV9 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
| SSV7 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
| SSV1 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
| ASV1 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
| SSV6 T8 Promoter       | TAGGCCCTTTTTAAAAAA | |
Figure 9a: 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.
Figure 9b: Conservation Among Non-Canonical Fusellovirus Promoters

| Tx Consensus Promoter | BRE | TATA Box |
|-----------------------|-----|----------|
| SSV10 Tx Promoter     | AAAAATTTTTTTTTAAGC |
| SSV7 Tx Promoter      | AAAAATTTTTTTTTTCC  |
| ASV1 Tx Promoter      | AAAAATTTTTTATTTAAAG |
| SSV2 Tx Promoter      | AAAAATTTTTTTTTTATTC |
| SSV8 Tx Promoter (ORF C82) | AAACTTTTTTTTTTATTC |
| SSV8 Tx Promoter (ORF B74) | AAAATTTTTTTTTTATTC |
| SSV1 Tx Promoter      | AAAATTTTTTTTTTATTTAC|
| SSV3 Tx Promoter      | AAGATATTATTAAAGC    |
| SSV4 Tx Promoter      | AAGATATTATTAAAGC    |
| SSV5 Tx Promoter      | AAGATATTATTAAAGC    |
| SSV9 Tx Promoter      | AAGATATTATTAAAGC    |

| Tint Consensus Promoter | BRE | TATA Box |
|-------------------------|-----|----------|
| SSV10 Tint Promoter    | TCATCCTTTTTTTTAA   |
| SSV5 Tint Promoter     | CACTCCTTTTTTTAAAC  |
| ASV1 Tint Promoter     | CAATGCCCTTTTTAAAC  |
| SSV6 Tint Promoter     | ATGAAATTATTATTC    |
| SSV2 Tint Promoter     | AATGAGATTATTAAAC   |
| SSV3 Tint Promoter     | AATGAGATTATTAAAC   |
| SSV4 Tint Promoter     | AATGAGATTATTAAAC   |
| SSV7 Tint Promoter     | AATGAGATTATTAAAC   |
| SSV9 Tint Promoter     | AATGAGATTATTAAAC   |

| T5' Consensus Promoter | BRE | TATA Box |
|------------------------|-----|----------|
| SSV10 T5' Promoter    | CATAGATATAATATAAG  |
| SSV8 T5' Promoter     | CATAGATATAATATAAG  |
| SSV2 T5' Promoter     | CATAGATATAATATAAG  |
| SSV4 T5' Promoter     | CATAGATATAATATAAG  |
| SSV3 T5' Promoter     | TGCAGTATAATATAAG   |
| SSV5 T5' Promoter     | TGCAGTATAATATAAG   |
| SSV6 T5' Promoter     | TGCAGTATAATATAAG   |

Figure 9b: 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.
Predicted Structures of SSV10 Proteins

Due to Fusellovirus gene products lack of homology with known proteins, primary sequence data is not always a sufficient source for comparison\(^\text{49}\). Although some primary protein sequences reveal conserved motifs, these archaeal viruses often deviate in sequence and codon usage from known proteins in bacteria and eukarya, and even from each other within the family. However, protein structure, particularly of viruses, is often conserved even in the absence of obvious sequence similarity\(^\text{96}\). In order to predict function, all SSV10 ORFs were translated and structural predictions (Table 5) were generated using the Phyre2 web portal for protein modeling\(^\text{42}\). 3D structural modeling using Phyre2\(^\text{42}\) makes it possible to compare these viral proteins to known protein structures in order to identify potential structures in viral proteins previously thought not to contain homologues either within or outside of the *Fuselloviridae*. SSV10 ORF B104 is a great example of this analysis. Phyre2\(^\text{42}\) structural predictions for ORF B104 indicate a high confidence model matching the structure of A100 from SSV1, however BLASTP\(^\text{2}\) searches with either ORF do not result in hits to the other. These and other findings utilizing Phyre2\(^\text{42}\) have expanded our understanding of the viral life cycle.

The 40 SSV10 ORF structural predictions were partitioned into three groups — low (≤44), middle (45-84), and high confidence (≥85) — based on the confidence rating generated by Phyre2\(^\text{42}\) for a significant portion (25%) of the query (Table 5). 12 of the 40 annotated ORFs in SSV10 returned no significant
structural matches of any kind. Of the matches, 12 were in the low confidence bracket, 4 were in the middle bracket, and 12 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 previously been characterized experimentally\(^{18,41,44,106}\). SSV10 ORFs F199, F64, E62, F73, B104, and C127 are all predicted matches to high resolution structures of homologous proteins from SSV1 (D63\(^{47}\), A100, and B129\(^{51}\)) or SSV8 (D212\(^{62}\) and E73\(^{91}\)). Additionally, uncharacterized SSV10 ORFs B65 and A81 encode a predicted copG-like and C2H2 zinc finger-like transcriptional regulators, respectively. SSV10 ORF D154 in the T5 transcript and C112 in the T8 transcript encode MarR-like and AsrR-like transcriptional regulators respectively. SSV10 ORF B205 encodes a Cas4-like protein product which has perfect confidence rating match over 95% of the sequence to a monomer of the *Sulfolobus solfataricus* Cas4 gene SSO0001\(^{52}\).

SSV10 ORFs A83, B94, D152, and D60 fall into the middle confidence bracket, with about \(\frac{1}{3}\) of their length matching predicted domains thought to be associated with the formation of a virus factory at the cellular membrane. SSV10 ORF A83 matches a centrosomin-like domain. ORF B94 shares folds similar to the DNA-binding domain of the lambda phage repressor. ORF D152 shares tertiary similarities with polymerase elongation factors and may be
distantly homologous to the archaeal transcription elongation factor Spt5. The insertion mutant generated in ORF D152 abrogated function of SSV10, which may be indicative of virus-specific interactions promoting effective transcription of viral mRNAs.

Twelve more ORFs make up the low confidence bracket, making predictions about function more challenging. However, two of these ORFs are of particular note: SSV10 ORFs E138 and C81. Interestingly, SSV10 ORFs E138 and C81 are predicted to be involved with deubiquitination and ubiquitin activation, respectively. Ubiquitin-like proteins are found in *Sulfolobus*.

Assuming a similar transcription cycle to SSV1, E138 would be expressed in the ‘early’ T5 transcript, where the deubiquitination of viral proteins could be vital to the viability of the virus during the infection cycle. Conversely, ORF C81 would be expressed in the ‘late’ transcript and could be responsible for the ubiquitination of host or viral proteins that may need to be degraded prior to egress from the host.

Finally, SSV10 encodes 11 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 (Figure 2). Since it would seem that fuselloviruses are forming virus factories at the cellular membrane, 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.

**Structural Predictions Provide Insight into the Life Cycle of the Fuselloviridae**

While members of the *Fuselloviridae* integrate into their hosts\(^1\), they do not establish a true lysogenic state in which the viral life cycle reaches a point of dormancy after integration. Rather, the *Fuselloviridae* maintain a replicative episome for sustained virus production even after integration into the host genome\(^1\). SSV1 is the only fusellovirus shown to be UV-inducible\(^2\), and the regulation of newly-described transcript T\(_{\text{lys}}\)\(^2\) seems to share an inverse relationship with this induced state and its transcript, T\(_{\text{ind}}\). Although the SSV1 gene F55 has been proposed to maintain this ‘carrier’ state\(^2\), both the T\(_{\text{lys}}\) and T\(_{\text{ind}}\) transcripts are not obviously maintained across the viral family, necessitating some other mechanism for the temporal expression of the viral genome and maintenance of the ‘carrier’ state\(^2\) in SSVs lacking an F55 homologue.

SSV10 ORFs E62, F73, and B65 are predicted to take on copG-like RHH folds similar to SSV1 ORFs F55 and C80 and share a subset of conserved residues between them. SSV10 ORFs E62 and B65 are encoded on the T5 and T6 transcripts, respectively, while SSV10 ORF F73 is encoded on the short T5’
transcript. Additionally, SSV10 ORF B94 is also of particular interest as it is predicted to undergo folding consistent with the DNA-binding domain of the lambda repressor (Table 5). These folds may implicate this ORF as another transcriptional regulator present in the genome, similar to copG-like ribbon-helix-helix (RHH) proteins found elsewhere in Fusellovirus genomes. Similar to the predicted RHH ORFs of SSV10, SSV10 ORF B94 is encoded early on in its transcript (T9). Moreover, SSV10 ORF B94 and its homologues are conserved across the family and may play a role in regulating transcription of the T9 transcript. It should be noted that the promoters for the T9 transcript are nearly identical to the T3 and T4/7/8 promoters (Figure 9a), so it may be that if SSV10 ORF B94, the last of these four to be expressed, is capable of regulating all three of these transcripts. If these four ORFs are indeed similar in function to the SSV1 ORF F55 it may be that they are binding to their own promoters, negatively regulating the expression of their respective transcripts.

Although SSV1 ORF F55 did not tolerate insertional mutation, which may have disrupted the putative origin of replication, a deletion of the SSV1 ORF F55 generated a functional mutant41. This deletion seemingly led to overproduction of virions per transmission electron micrographs and plaque assay (Personal communication). However, a transposon insertion in F73 of SSV10 generated a functional mutant, and the halos of these infected cultures exhibited a slightly larger zone of inhibition than the positive control, which may be indicative of increased virus production.
Structures of products of ORFs F93, D63, F112, and B129 from SSV1 and ORFs E73 and D212 from SSV8 have been solved\textsuperscript{47,48,62,63,91}. All except SSV8 ORF D212, which encodes an incompletely-conserved predicted viral nuclease, have been implicated in the regulation of gene expression during the infection cycle, as they contain canonical DNA-binding domains. SSV1 ORF B129 encodes a C2H2 zinc finger while the other four are helix-turn-helix proteins. SSV1 ORFs F93, D63, B129, and SSV8 ORF E73 all form homodimers\textsuperscript{47,48,91}, while SSV1 ORF F112 is thought to exist as a monomer\textsuperscript{63}. Apart from ORF F112, which is thought to be unique to SSV1, homologues of these putative transcriptional regulators can be found in SSV10 as well. Although little is known about the specific binding interactions of these regulators, with the exception of the predicted viral nuclease it is likely that they are responsible for regulated gene expression throughout the infection cycle.

SSV10 ORFs C82, B94, C281, C102, and A110 — all of which contain at least one putative transmembrane helix — are encoded on the putative ‘early’ T6 and T9 transcripts. The conserved SSV10 ORF C250, a DnaA-like AAA+ ATPase which has previously been implicated in viral replication is also encoded on the T9 transcript. Intriguingly, structural predictions of the T9-encoded SSV10 ORF A83 indicate that the gene product may share some functional characteristics with eukaryotic centrosomin-like proteins (Table 5) which are responsible for the organization of centrosomes in dividing Drosophila cells\textsuperscript{61}. 
Thus, SSV10 ORF A83 may aid in the formation of a viral replication focus through interactions with the cytoskeleton of the host.

Recently, and for the first time in Archaea, viral replication foci were shown to be formed by the rudivirus *Sulfolobus islandicus* Rod-shaped Virus 2 (SIRV2)\(^60\). These viral “factories” are commonplace in both eukarya and bacteria, and are accompanied by sweeping reorganizations of the membrane, and in some cases the cytoskeletal structure, of the host cell\(^93\). Infection of *Sulfolobus* by SIRV2 was found to lead to the formation of viral factories, located around the edges of the host cells, and possibly even directly interacting with the cellular membrane\(^60\). Furthermore, SSV1 assembly and egress has been shown by electron tomography to occur at the cellular membrane in a fashion similar to budding by some eukaryotic viruses\(^77\). The predicted transmembrane proteins occurring in the putative ‘early’ T6 and T9 transcripts (Table 5) of SSV10 may be involved in establishing such a structure by appropriating and attaching to the host membrane. A second suite of transmembrane proteins are encoded on the late transcripts T3, Tx, and T4/7/8, as every ORF in these transcripts contains at least 1 transmembrane helix. These are likely to localize at the site of virus replication for effective genome packaging and eventual egress of the mature virion.
Table 5: Phyre2 Protein Structural Predictions for ORFs of SSV10

| ORF* (Transcript) | Phyre2 Prediction | Coverage/Confidence | Predicted Function |
|-------------------|-------------------|---------------------|--------------------|
| **C82 (T9)**      | Transmembrane Protein | —†                 | Replication Complex |
|                   |                   |                     | (2 TMH³)           |
| **A83 (T9)**      | Centrosomin¹      | 34%@55              | Replication Complex |
|                   | (PDB: 5I7C)       |                     |                    |
| **B94 (T9)**      | Lambda Repressor-like DNA-binding domain | 29%@45 | Replication Complex |
|                   |                   |                     | (3 TMH)            |
| **C281 (T9)**     | Transmembrane Protein | —                  | Replication Complex |
|                   |                   |                     | (1 TMH)            |
| **A89 (T9)**      | Matrix Protein    | 51%@23              | Replication Complex |
| **B150 (T9)**     | Type II Topoisomerase/Gyrase | 28%@30 | Replication Complex |
| **C250 (T9)**     | DnaA-like AAA+ ATPase⁴⁴ | 75%@>90 | Replication Initiation |
| **C64 (T9)**      | Phosphoglycerate mutase | 48%@36 | —                   |
| **D335 (Tint)**   | Tyrosine recombinase²¹ | 95%@100 | Viral Integration |
|                   | (PDB: 3UXU)       |                     |                    |
| **A35 (Tori)**    | —                 | —                   | —                  |
| **C34 (Tori)**    | DNA Binding       | 26%@16              | —                  |
| **A49 (Tori)**    | Transmembrane Protein | —                 | Replication Complex/origin |
|                   |                   |                     | (1 TMH)            |
| **E52 (T5)**      | Sigma2-like domain⁹ | 50%@11              | Transcriptional Regulator |
| **F199 (T5)**     | SSV8 ORF D212⁶² (PDB: 2W8M) | 49%@>90 | Viral Nuclease |
| **E138 (T5)**     | Deubiquitinating domain | 25%@38 | Degradation Control |
| Gene   | Description                                                                 | Annotation | Notes                  |
|--------|------------------------------------------------------------------------------|------------|------------------------|
| F64 (T5) | SSV1 ORF D63<sup>17</sup>                                                      | 92%@100    | ROP-like regulator     |
|        | (PDB: 1SKV)                                                                  |            |                        |
| D152 (T5) | Polymerase II Elongation Factor ell2                                          | 33%@53     | Replication Complex    |
|        | (PDB: 2E5N)                                                                  |            |                        |
| F111 (T5) | Transmembrane Protein                                                         | —          | Replication Complex    |
|        |                                                                            |            | (2 TMH)                |
| D154 (T5) | MarR-like Transcriptional Regulator<sup>108</sup>                            | 72%@>90    | Host Interference      |
|        | (PDB: 3F3X)                                                                  |            |                        |
| E62 (T5)  | SSV8 ORF E73<sup>31</sup>                                                     | 57%@44     | CopG-like RHH Regulator|
|        | (PDBe: 4aa1i)                                                                |            |                        |
| D60 (T5') | Transport Protein                                                             | 28%@46     | Replication Complex    |
| F73 (T5') | SSV8 ORF E73<sup>31</sup>                                                     | 98%@100    | CopG-like RHH Regulator|
|        | (PDBe: 4aa1i)                                                                |            |                        |
| E56     | Signaling Protein                                                             | 44%@31     | —                      |
| B104 (T6) | SSV1 ORF A100                                                                | 61%@87.4   | DNA binding scaffold   |
|        | (PDBe: 4lid)                                                                 |            |                        |
| C117 (T6) | Vaccinia virus protein B14<sup>14,15</sup>                                    | 38%@38.2   | Virulence Factor       |
|        | (PDB: 2VVY)                                                                  |            |                        |
| C59 (T6)  | CheY-like Protein                                                             | 80%@28     | Motility Inhibition    |
| B65 (T6)  | RHH Protein                                                                  | 66%@>90    | CopG-like RHH Regulator|
| A81 (T6)  | Eukaryal C2H2 Zinc Finger                                                     | 94%@>90    | Transcriptional Regulator|
| C102 (T6) | RNA editing uridylyl transferase                                              | 29%@36     | —                      |
| B205 (T6) | Crystal structure of SSO0001                                                  | 95%@100    | Anti-CRISPR/DNA Repair |
| C127 (T6) | SSV1 ORF B129 C2H2 Zinc Finger<sup>51</sup>                                  | 99%@100    | Transcriptional Regulator|

54
| Protein | Type | Function | TMH | Notes |
|---------|------|----------|------|-------|
| A110 (T6) | Transmembrane Protein | Virion Assembly/Egress | 3 | (PDB: 2WBT) |
| B298 (T3) | — | Virion Assembly/Egress | 5 |
| A74 (Tx) | Transmembrane Protein | Virion Assembly/Egress | 1 |
| A814 (VP4) (T4/7/8) | — | Terminal Fiber | — |
| C81 (T4/7/8) | Ubiquitin-activating enzyme E1 | Virion Assembly/Egress | 2 |
| A166 (T4/7/8) | Transmembrane Protein | Virion Assembly/Egress | 4 |
| C112 (T4/7/8) | ArsR Family | Host Interference | 1 |
| A138 (VP1) (T4/7/8) | — | Major Capsid Protein | — |
| C95 (VP3) (T4/7/8) | — | Minor Capsid Protein | — |

*Colors correlate to Figure 2
†No significant prediction
* Transmembrane Helices
Chapter 4: Summary & Future Directions

This work characterizes *Sulfolobus* Spindle-shaped virus 10 as a novel member of the *Fuselloviridae* family. Both its genetic and genomic characteristics are comparable to other members of the viral family, most notably SSV8. SSV10 is also shown to be genetically tractable, withstanding both insertion and deletion mutations in a fashion consistent with previous work done with SSV1\(^{40,41}\). Furthermore, SSV10 provides a necessary contrasting point of comparison, as its genome shares characteristics of and homology with members of the *Fuselloviridae* that are absent in SSV1. Although ORFs of the putative T5’ and T5 transcripts of SSV10 generally tolerate mutation, overall gene conservation in this region of SSV10 is greater than in SSV1, where a majority of genes are unique to SSV1. The more conserved genes in this region of SSV10 — ORFs E62, D154, F111, and D152 — are not homologous to genes found in SSV1 and do not tolerate insertion mutations, highlighting both genetic and potential life cycle differences between the two viruses. Also, in contrast to SSV1, mutations in the SSV10 integrase gene lead to a complete loss of infectivity irrespective of host. Thus, these findings, as well as continued comparative studies between SSV1 and SSV10, provide a broader understanding of the genetic requirements of members of the crenarchaeal virus family *Fuselloviridae*.

Similar to SSV1, however, SSV10 tolerates deletion mutations of the minor capsid gene VP3 leading to a similarly aberrant, elongated morphology of the virions produced. Recent structural studies of both *Aeropyrum pernix*
bacilliform virus 1 (APBV1)\textsuperscript{76} and \textit{Acidianus} tailed spindle virus (ATSV)\textsuperscript{38} propose intriguing capsid models that may provide insight into the morphological changes due to minor capsid protein deletions in fuselloviruses. Both ATSV and APBV1 utilize a single major capsid protein (MCP) whose quaternary structure is centered around interactions between mostly hydrophobic side chains, a quality shared by the major and minor capsid proteins in SSVs. Structural models of SSV10 major and minor capsid genes VP1 and VP3 generated using Phyre2\textsuperscript{42} are, when paired together, structurally homologous to the antiparallel four helix bundle structure of the ATSV MCP. Conversely, APBV1 encodes a MCP roughly half the size of ATSV, consisting only of an $\alpha$-$\beta$-$\beta$-$\alpha$ motif equivalent to the N-terminal half of the ATSV MCP, which leads to a bacilliform virion. This bacilliform structure is relevant as it pertains to VP3 deletions in SSV1 and SSV10. The wild-type morphology of SSV1 and SSV10 is similar to the fusiform structure of ATSV, and their capsid components are predicted to be structurally similar. However, when the minor capsid gene VP3 is deleted from SSV1 and SSV10, they are left with only VP1, which is structurally similar to the MCP of APBV1. In the absence of VP3, both SSV1 and SSV10 exhibit a bacilliform morphology similar to that of APBV1. It is tempting, then, to hypothesize that wild-type SSVs are fusiforms assembled similarly to ATSV, and deletion mutants lacking VP3 lead to bacilliform virions assembled similarly to APBV1. Unfortunately, the high-resolution structural nature of fusellovirus capsid genes is unknown, and whole particle reconstructions of SSV1\textsuperscript{98} are too low-resolution
to determine the nature of its assembly as it pertains to interactions and concentration of VP1 and VP3. Attempts to express wild-type VP1, as well as point mutants of VP1 are currently underway, and will hopefully generate the first high-resolution structures of a fusellovirus capsid protein.

SSV10 encodes a Cas4-like gene — ORF B205 — deletions of which lead to a loss of infectivity over time, a phenotype indicative of a CRISPR/Cas response to infection from *S. solfataricus*[^58], and not observed with other insertion or deletion mutants (Figure 5). 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. Although no structure has yet been obtained from these virus-encoded Cas4-like proteins, sequence alignments to the host-encoded Cas4 SSO0001[^52] indicate they maintain all of the conserved residues required for endonuclease and 5’ → 3’ exonuclease activities, as well as the formation of higher order structures[^52] (Figure 4). SSV10 ORF B205 has homologues in 7 other SSVs, but is not found in SSV1, which again highlights significant genetic differences between these two model fuselloviruses that implicate a divergence in their interactions with their host organism.

In recent years, an increase in studies of virally-encoded CRISPR/Cas systems have begun to elucidate the complicated genetic arms-race occurring between viruses and their hosts. One such study presents intriguing data that point to the potential mechanisms of these virally-encoded Cas4 genes[^27]. *S.*
solfataricus cells infected with SSV2, which encodes a homologue of SSV10 ORF B205, were shown to delete CRISPR repeats and spacers from the leader-proximal end of one of its CRISPR loci. It is hypothesized that the integration of SSV2 into the host chromosome drives these deletions, which assures that the host CRISPR/Cas response does not attack the integrated provirus. Furthermore, this phenotype is not observed in Sulfolobus infected with SSV1 (which lacks a Cas4-like homologue), nor is the overall CRISPR/Cas response from the cell to SSV1 infection similar to that observed when cells are infected by SSV2. However, the study does not acknowledge the presence of the virally-encoded Cas4-like gene in SSV2 or the potential role it may play. Another study published the same year\textsuperscript{39} presents an interesting alternative mechanism for virally-encoded Cas4 genes. Bacteriophages that infect Camplyobacter jejuni also encode a Cas4 gene homologous to those found in SSVs and S. solfataricus. Rather than deleting portions of its CRISPR loci similar to SSV2-infected S. solfataricus, C. jejuni infected with phage containing Cas4 homologs were shown to incorporate host-derived spacers into its only CRISPR locus. After multiple passages of infected cultures, newly acquired spacers present in the C. jejuni CRISPR locus could be mapped back to the host chromosome and did not match any portion of the genome of the infecting phages. In light of these data, the SSV10 ORF B205 deletion mutant presented in this work provides a new and unique opportunity to explore the host response to
infection. It will be fascinating to see whether the presence or absence of a virally-encoded Cas4 gene leads to either or both of these host responses.

This work also presents a comprehensive comparison of transcription promoter elements across the *Fuselloviridae*, highlighting the similarities and differences that may play a role in the regulation of expression of viral transcripts. This is epitomized by the presence of multiple ORFs — E62, B65, and F73 in SSV10 — encoding predicted ribbon-helix-helix DNA binding proteins on multiple transcripts in fusellovirus genomes which are likely to play a role in negative regulation of their own transcripts. Although the specific mechanisms of these gene products are unclear as it relates to the temporal regulation of gene expression observed during fusellovirus infections\textsuperscript{26,27,86}, they do share putative structural and sequence similarity with the SSV1 ORF F55, which has tentatively been shown to bind to its own promoter, as well as the promoter elements of the T5, T6, and T\textsubscript{ind} transcripts of SSV1\textsuperscript{28,29}. SSV1 ORF F55 has been implicated in maintaining the ‘carrier’ or quasi-lysogenic state observed \~8-10 hours after a UV-induced SSV1 infection of *S. solfataricus* by negatively regulating the expression of the ‘early’ T5 and T6 transcripts. Moreover, similar expression of SSV2 ORF D79, a homologue of SSV10 ORF F73, was observed in non-induced, stably infected *S. solfataricus*\textsuperscript{27}, implicating these homologues may be functionally similar to SSV1 ORF F55. However, it is likely that the other RHH DNA-binding genes present in these genomes are responsible for the specific regulation of transcription once the replicative infection process of
fuselloviruses has been fully induced. One way to probe the connection between these ORFs would be to quantify virus production in mutants of SSV10 ORFs F73 and/or B65 via plaque assay and qPCR. RNA-seq experiments should also be performed on *S. solfataricus* infected with wild-type SSV10 compared to SSV10 with mutations in these ORFs in order to generate whole-genome transcript data for viral expression, the role these ORFs play in amount of transcript, as well as alterations to host gene regulation upon infection with SSV10.

Lastly, an extensive analysis of structural predictions generated for every ORF in the SSV10 genome has been presented. These predictions have allowed putative functions to be assigned to nearly every potential gene product encoded by SSV10, which will hopefully spur more directed experimentation and research into the life cycle and virus-host interactions of the *Fuselloviridae*. For many of these ORFs it is the first time that these functions have been predicted in this system, and most of the predicted activities are involved in interactions with, or the outright takeover of, host functions. Often implicated in virus infections are changes to the host replication, transcription, and translation machinery which are necessary for virus propagation. However, these predictions extend far beyond these processes, with viral gene products potentially regulating cell stress response (*i.e. MarR* and *CRISPR/Cas*), ubiquitination pathways, and motility. Again, various SSV10 transposon insertion mutants generated in this study will be vital in probing the accuracy of
these predictions and may greatly expand our understanding of the virus-host interactions in this system. Overall, the genetic tractability of *Sulfolobus* spindle-shaped virus 10 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.
References

1. Alekshun MN, Levy SB. Regulation of chromosomally mediated multiple antibiotic resistance: The mar regulon. 1997;41(10):2067–2075.
2. Altschul SF et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. 1997;25(17):3389–3402.
3. Barry ER, Bell SD. DNA Replication in the Archaea. 2006;70(4):876–887. doi:10.1128/MMBR.00029-06
4. Basehoar AD, Zanton SJ, Pugh BF. Identification and distinct regulation of yeast TATA box-containing genes. 2004;116(5):699–709. doi:10.1016/S0092-8674(04)00205-3
5. Bautista MA, Zhang C, Whitaker RJ. Virus-induced dormancy in the archaeon Sulfolobus islandicus. 2015;6(2):1–8. doi:10.1128/mBio.02565-14
6. Bize A, Karlsson EA, Ekefjard K, Quax TE, Pina M, Prevost MC, Forterre P, Tenailleon O, Bernarder R, Prangishvili D. A unique virus release mechanism in the Archaea. 2009;106(27):11306–11311. doi:10.1073/pnas.0901238106
7. Bondy-Denomy, Joe; Pawluk, April; Maxwell, Karen L.; Davidson AR. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. 2013;1(Part III):429–432. doi:10.1038/nature11723
8. Bondy-Denomy J, Garcia B, Strum S, Du M, Maryclare F, Hidalgo-reyes Y, Wiedenheft B, Maxwell KL, Alan R. Multiple mechanisms for CRISPR–Cas inhibition by anti–CRISPR proteins. 2016;526(7571):136–139. doi:10.1038/nature15254
9. Burton SP, Burton ZF. The σ enigma: Bacterial σ factors, archaeal TFB and eukaryotic TFIIB are homologs. 2014;5(4):e967599. doi:10.4161/21541264.2014.967599
10. Cannio R, Cotursi P, Rossi M, Bartolucci S. An autonomously replicating transformation vector for Sulfolobus solfataricus. 1998;180(12):3237–3240.
11. Ceballos RM, Marceau CD, Marceau JO, Morris S, Clore AJ, Stedman KM. Differential virus host-ranges of the Fuselloviridae of hyperthermophilic Archaea: Implications for evolution in extreme environments. 2012;3(AUG):1–10. doi:10.3389/fmicb.2012.00295
12. Chan PP, Holmes AD, Smith AM, Tran D, Lowe TM. The UCSC Archaeal Genome Browser: 2012 update. 2012;40(D1):646–652. doi:10.1093/nar/gkr990
13. Chen L, Brügger K, Skovgaard M, She Q, Torarinsson E, Greve B, Zibat A, Klenk H-P, Garrett RA, Bru K, et al. The Genome of Sulfolobus acidocaldarius, a Model Organism of the Crenarchaeota. 2005;187(14):4992–4999. doi:10.1128/JB.187.14.4992
14. Chen RAJ, Jacobs N, Smith GL. Vaccinia virus strain Western Reserve protein B14 is an intracellular virulence factor. 2006;87(6):1451–1458. doi:10.1099/vir.0.81736-0
15. Chen RAJ, Ryzhakov G, Cooray S, Randow F, Smith GL. Inhibition of IkB kinase by vaccinia virus virulence factor B14. 2008;4(2). doi:10.1371/journal.ppat.0040022
16. Chowdhury S, Carter J, Rollins MCF, Golden SM, Jackson RN, Hoffmann C,
Nosaka L, Bondy-Denomy J, Maxwell KL, Davidson AR, et al. Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. 2017;169(1):47–57.e11. doi:10.1016/j.cell.2017.03.012

17. Clore AJ. The Family Fuselloviridae: Diversity and replication of a hyperthermic virus infecting the Archaeon genus Sulfolobus. 2008;Portland S.

18. Clore AJ, Stedman KM. The SSV1 viral integrase is not essential. 2007;361(1):103–111. doi:10.1016/j.virol.2006.11.003

19. Dellas N, Snyder JC, Bolduc B, Young MJ. Archaeal Viruses: Diversity, Replication, and Structure. 2014. doi:10.1146/annurev-virology-031413-085357

20. Duggin IG, Bell SD. The chromosome replication machinery of the archaeon Sulfolobus solfataricus. 2006;281(22):15029–15032. doi:10.1074/jbc.R500029200

21. Eilers BJ, Young MJ, Lawrence CM. The Structure of an Archaeal Viral Integrase Reveals an Evolutionarily Conserved Catalytic Core yet Supports a Mechanism of DNA Cleavage in trans. 2012;86(15):8309–8313. doi:10.1128/JVI.00547-12

22. Erdmann S, Le Moine Bauer S, Garrett RA. Inter-viral conflicts that exploit host CRISPR immune systems of Sulfolobus. 2014;91(5):900–917. doi:10.1111/mmi.12503

23. Di Fiore A, Fiorentino G, Vitale RM, Ronca R, Amodeo P, Pedone C, Bartolucci S, De Simone G. Structural Analysis of BldR from Sulfolobus solfataricus Provides Insights into the Molecular Basis of Transcriptional Activation in Archaea by MarR Family Proteins. 2009;388(3):559–569. doi:10.1016/j.jmb.2009.03.030

24. Fiorentino G, Del Giudice I, Bartolucci S, Durante L, Martino L, Del Vecchio P. Identification and physicochemical characterization of BldR2 from Sulfolobus solfataricus, a novel archaeal member of the MarR transcription factor family. 2011;50(31):6607–6621. doi:10.1021/bi200187j

25. Fröls S, Gordon PMK, Panlilio MA, Duggin IG, Bell SD, Sensen CW, Schleper C. Response of the hyperthermophilic archaeon Sulfolobus solfataricus to UV damage. 2007;189(23):8708–8718. doi:10.1128/JB.01016-07

26. Fröls S, Gordon PMK, Panlilio MA, Schleper C, Sensen CW. Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. 2007;365(1):48–59. doi:10.1016/j.virol.2007.03.033

27. Fusco S, Liguori R, Limauro D, Bartolucci S, She Q, Contursi P. Transcriptome analysis of Sulfolobus solfataricus infected with two related fuselloviruses reveals novel insights into the regulation of CRISPR-Cas system. 2015;118(April):322–332. doi:10.1016/j.biochi.2015.04.006

28. Fusco S, She Q, Bartolucci S, Contursi P. 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. 2013;87(10):5926–36. doi:10.1128/JVI.00458-13

29. Fusco S, She Q, Fiorentino G, Bartolucci S, Contursi P. Unravelling the Role of the F55 Regulator in the Transition from Lysogeny to UV Induction of
Sulfolobus Spindle-Shaped Virus 1. 2015;89(12):6453–6461. doi:10.1128/jvi.00363-15

30. Garrett R, Shah S, Erdmann S, Liu G, Mousaei M, León-Sobrino C, Peng W, Gudbergsdottir S, Deng L, Vestergaard G, et al. CRISPR-Cas Adaptive Immune Systems of the Sulfolobales: Unravelling Their Complexity and Diversity. 2015;5(1):783–817. doi:10.3390/life5010783

31. Garrett RA, Vestergaard G, Shah SA. Archaeal CRISPR-based immune systems: Exchangeable functional modules. 2011;19(11):549–556. doi:10.1016/j.tim.2011.08.002

32. Goulet A, Spinelli S, Blangy S, Van Tilbeurgh H, Leulliot N, Basta T, Prangishvili D, Cambillau C, Campanacci V. The crystal structure of ORF14 from Sulfolobus islandicus filamentous virus. 2009;76(4):1020–1022. doi:10.1002/prot.22448

33. Green MR, Sambrook J. Molecular Cloning: A Laboratory Manual (4th Edition)

34. Gudbergsdottir S, Deng L, Chen Z, Jensen JVK, Jensen LR, She Q, Garrett RA. Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. 2011;79(1):35–49. doi:10.1111/j.1365-2958.2010.07452.x

35. He F, Bhoobalan-Chitty Y, Van LB, Kjeldsen AL, Dedola M, Makarova KS, Koonin EV, Brodersen DE, Peng X. Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. 2018;3(4):461–469. doi:10.1038/s41564-018-0120-z

36. He F, Vestergaard G, Peng W, She Q, Peng X. CRISPR-Cas type I-A Cascade complex couples viral infection surveillance to host transcriptional regulation in the dependence of Csa3b. 2016;45(4):gkw1265. doi:10.1093/nar/gkw1265

37. Held NL, Whitaker RJ. Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. 2009;11(2):457–466. doi:10.1111/j.1462-2920.2008.01784.x

38. Hochstein R, Bollschweiler D, Dharmavaram S, Lintner NG, Plitzko JM, Bruinsma R, Engelhardt H, Young MJ, Klug WS, Lawrence CM. Structural studies of Acidianus tailed spindle virus reveal a structural paradigm used in the assembly of spindle-shaped viruses. 2018;201719180. doi:10.1073/pnas.1719180115

39. Hooton SPT, Connerton IF. Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. 2015;6(JAN):1–9. doi:10.3389/fmicb.2014.00744

40. Iverson E, Stedman K. A genetic study of SSV1, the prototypical fusellovirus. 2012;3(JUN):200. doi:10.3389/fmicb.2012.00200

41. Iverson EA, Goodman DA, Gorchels ME, Stedman KM. 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. 2017;91(10):e02406-16. doi:10.1128/JVI.02406-16

42. Kelley L a, Mezulis S, Yates CM, Wass MN, Sternberg MJE. Europe PMC
Funders Group The Phyre2 web portal for protein modelling, prediction and analysis. 2015;10(6):845–858. doi:10.1038/nprot.2015.053.
The
43. Kent WJ. BLAT — The BLAST-Like Alignment Tool. 2002;12:656–664. doi:10.1101/gr.229202.
44. Koonin E V. Archaebacterial virus SSV1 encodes a putative DNA-binding protein. 1992;20(5):1143. doi:10.1093/nar/20.5.1143
45. Kosa PF, Ghosh G, Dedecker BS, Sigler PB. Notes: The 2.1 Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein—transcription factor (II) B. 1997;94(ii):6042–6047. doi:10.1073/pnas.94.12.6042
46. Kraft P, Gauss GH, Young M, Lawrence CM. Structural Studies of Crenarchaeal Viral Proteins: Structure Suggests Function. 2005;(June):305–316.
47. Kraft P, Kümmel D, Oeckinghaus A, Gauss GH, Wiedenheft B, Young M, Lawrence CM. Structure of D-63 from sulfolobus spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. 2004;78(14):7438–42. doi:10.1128/JVI.78.14.7438-7442.2004
48. Kraft P, Oeckinghaus A, Kümmel D, Gauss GH, Gilmore J, Wiedenheft B, Young M, Lawrence CM. Crystal Structure of F-93 from Sulfolobus Spindle-Shaped Virus 1, a Winged-Helix DNA Binding Protein. 2004;78(21):11544–11550. doi:10.1128/JVI.78.21.11544
49. Krupovic M, Cvirkaite-Krupovic V, Ionzo J, Prangishvili D, Koonin E V. Viruses of archaea: Structural, functional, environmental and evolutionary genomics. 2018;244(July 2017):181–193. doi:10.1016/j.virusres.2017.11.025
50. Krupovic M, Cvirkaite-Krupovic V, Prangishvili D, Koonin E V. Evolution of an archaeal virus nucleocapsid protein from the CRISPR-associated Cas4 nuclease. 2015;10(1):65. doi:10.1186/s13062-015-0093-2
51. Lawrence CM, Menon S, Eilers BJ, Bothner B, Khayat R, Douglas T, Young MJ. Structural and functional studies of archaeal viruses. 2009;284(19):12599–12603. doi:10.1074/jbc.R800078200
52. Lemak S, Beloglazova N, Nocek B, Skarina T, Flick R, Brown G, Popovic A, Joachimiak A, Savchenko A, Yakunin AF. Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from sulfolobus solfataricus. 2013;135(46):17476–17487. doi:10.1021/ja408729b
53. Liu G, She Q, Garrett RA. Diverse CRISPR-Cas responses and dramatic cellular DNA changes and cell death in pKEF9-conjugated Sulfolobus species. 2016;44(9):4233–4242. doi:10.1093/nar/gkw286
54. Liu T, Liu Z, Ye Q, Pan S, Wang X, Li Y, Peng W, Liang Y, She Q, Peng N. Coupling transcriptional activation of CRISPR-Cas system and DNA repair genes by Csa3a in Sulfolobus islandicus. 2017;45(15):8978–8992. doi:10.1093/nar/gkx612
55. Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R. Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. 2004;101(18):7046–51. doi:10.1073/pnas.0400656101
56. Makarova KS, Koonin E V. Archaeal ubiquitin-like proteins: Functional
versatility and putative ancestral involvement in tRNA modification revealed by comparative genomic analysis. 2010;2010:9–13. doi:10.1155/2010/710303
57. Manica A, Zebec Z, Steinkellner J, Schleper C. Unexpectedly broad target recognition of the CRISPR-mediated virus defence system in the archaeon Sulfolobus solfataricus. 2013;41(22):10509–10517. doi:10.1093/nar/gkt767
58. Manica A, Zebec Z, Teichmann D, Schleper C. In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaean. 2011;80(2):481–491. doi:10.1111/j.1365-2958.2011.07586.x
59. Martin A, Yeats S, Janekovic D, Reiter WD, Aicher W, Zillig W. SAV 1, a temperate u.v.-inducible DNA virus-like particle from the archaeabacterium Sulfolobus acidocaldarius isolate B12. 1984;3(9):2165–8.
60. Martinez-Alvarez, L; Deng, Ling; Peng X. Formation of a Viral Replication Focus in Sulfolobus Cells Infected by the Rudivirus Sulfolobus islandicus Rod-Shaped Virus 2. 2017;91(13):1–12.
61. Megraw TL, Li K, Kao LR, Kaufman TC. The centrosomin protein is required for centrosome assembly and function during cleavage in Drosophila. 1999;126(13):2829–39.
62. Menon SK, Eilers BJ, Young MJ, Lawrence CM. The crystal structure of D212 from sulfolobus spindle-shaped virus ragged hills reveals a new member of the PD-(D/E)XK nuclease superfamily. 2010;84(12):5890–7. doi:10.1128/JVI.01663-09
63. Menon SK, Maaty WS, Corn GJ, Kwok SC, Eilers BJ, Kraft P, Gillitzer E, Young MJ, Bothner B, Lawrence CM. Cysteine usage in Sulfolobus spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. 2008;376(2):270–278. doi:10.1016/j.virol.2008.03.026
64. Muskheishvili G, Palm P, Zillig W. SSV1-encoded site-specific recombination system in Sulfolobus shibatae. 1993;237(3):334–342. doi:10.1007/BF00279436
65. Nakagawa S, Niimura Y, Gojobori T. Comparative genomic analysis of translation initiation mechanisms for genes lacking the Shine–Dalgarno sequence in prokaryotes. 2017;45(7):3922–3931. doi:10.1093/nar/gkx124
66. Palm P, Schleper C, Grampp B, Yeats S, McWilliam P, Reiter WD, Zillig W. Complete nucleotide sequence of the virus SSV1 of the archaeabacterium Sulfolobus shibatae. 1991;185(1):242–250.
67. Pawluk A, Bondy-Denomy J, Cheung VHW, Maxwell KL, Davidson AR. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of pseudomonas aeruginosa. 2014;5(2):1–7. doi:10.1128/mBio.00896-14
68. Pawluk A, Staals RHJ, Taylor C, Watson BNJ, Saha S, Fineran PC, Maxwell KL, Davidson AR. Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. 2016;1(8):16085. doi:10.1038/nmicrobiol.2016.85
69. Peng N, Deng L, Mei Y, Jiang D, Hu Y, Awayez M, Liang Y, She Q. A Synthetic Arabinose-Inducible Promoter Confers High Levels of Recombinant Protein Expression in Hyperthermophilic Archaeon Sulfolobus islandicus. 2012. doi:10.1128/AEM.00855-12
70. Peng X. 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. 2008;154(2):383–391. doi:10.1099/mic.0.2007/012963-0

71. Peng X, Garrett RA, She Q. Archaeal viruses—novel, diverse and enigmatic. 2012;55(5):422–433. doi:10.1007/s11427-012-4325-8

72. Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax. 2012;194(10):2491–2500. doi:10.1128/JB.00206-12

73. Prangishvili D. The Wonderful World of Archaeal Viruses. 2013. doi:10.1146/annurev-micro-092412-155633

74. Prangishvili D, Forterre P, Garrett RA. Viruses of the Archaea: a unifying view. 2006;4(11):837–848. doi:10.1038/nrmicro1527

75. Prangishvili D, Krupovic M. A new proposed taxon for double-stranded DNA viruses, the order “Ligamenvirales.” 2012;157(4):791–795. doi:10.1007/s00705-012-1229-7

76. Ptchelkine D, Gillum A, Mochizuki T, Lucas-Staat S, Liu Y, Krupovic M, Phillips SEV, Prangishvili D, Huiskonen JT. Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. 2017;8(1):7–12. doi:10.1038/s41467-017-01668-0

77. Quemin ERJ, Chlanda P, Sachse M, Forterre P, Prangishvili D, Krupovic M. Eukaryotic-like virus budding in archaea. 2016;7(5):3–7. doi:10.1128/mBio.01439-16

78. Quemin ERJ, Lucas S, Daum B, Quax TEF, Kühlbrandt W, Forterre P, Albers S-V, Prangishvili D, Krupovic M. First insights into the entry process of hyperthermophilic archaeal viruses. 2013;87(24):13379–85. doi:10.1128/JVI.02742-13

79. Quemin ERJ, Pietilä MK, Oksanen HM, Forterre P, Rijpstra WIC, Schouten S, Bamford DH, Prangishvili D, Krupovic M. Sulfolobus Spindle-Shaped Virus 1 Contains Glycosylated Capsid Proteins, a Cellular Chromatin Protein, and Host-Derived Lipids. 2015;89(22):11681–11691. doi:10.1128/JVI.02270-15

80. Quemin ERJ, Quax TEF. Archaeal viruses at the cell envelope: Entry and egress. 2015;6(JUN):552. doi:10.3389/fmicb.2015.00552

81. Qureshi SA, Jackson SP. Sequence-specific DNA binding by the S. shibatae TFIIB homolog, TFB, and its effect on promoter strength. 1998;1(3):389–400. doi:10.1016/S1097-2765(00)80039-8

82. Redder P, Peng X, Brügger K, Shah SA, Roesch F, Greve B, She Q, Schleper C, Forterre P, Garrett RA, et al. Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interval recombination mechanism. 2009;11(11):2849–2862. doi:10.1111/j.1462-2920.2009.02009.x

83. Reiter W-D, Palm P, Zillig W. Analysis of Transcription in the Archaebacterium Sulfolobus indicates that archaeobacterial promoters are homologous to eukaryotic pol II promoters. 1998;16. doi:10.1093/nar/gkn907
84. Reiter WD, Palm P, Yeats S. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. 1989;17(5):1907–1914. doi:10.1093/nar/gkn907
85. Reiter WD, Palm P, Yeats S, Zillig W. Gene expression in archaebacteria: Physical mapping of constitutive and UV-inducible transcripts from the Sulfolobus virus-like particle SSV1. 1987;209(2):270–275. doi:10.1007/BF00329653
86. Ren Y, She Q, Huang L. Transcriptomic analysis of the SSV2 infection of Sulfolobus solfataricus with and without the integrative plasmid pSSVi. 2013;441(2):126–134. doi:10.1016/j.viroi.2013.03.012
87. Rensen EI, Mochizuki T, Quemin E, Schouten S, Krupovic M, Prangishvili D. A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. 2016;113(9):2478–2483. doi:10.1073/pnas.1518929113
88. Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD. Identification of Two Origins of Replication in the Single Chromosome of the Archaeon Sulfolobus solfataricus. 2004;116(1):25–38. doi:10.1016/S0092-8674(03)01034-1
89. Romero a, García P. Initiation of translation at AUC, AUA and AUU codons in Escherichia coli. 1991;68(3):325–30.
90. Samson RY, Xu Y, Gadelha C, Stone TA, Faqiri JN, Li D, Qin N, Pu F, Liang YX, She Q, et al. Specificity and Function of Archaeal DNA Replication Initiator Proteins. 2013;3(2):485–496. doi:10.1016/j.celrep.2013.01.002
91. Schlenker C, Goel A, Tripet BP, Menon S, Willi T, Dlakic M, Young MJ, Lawrence CM, Copie V. Structural studies of E73 from a hyperthermophilic archaeal virus identify the “RH3” domain, an elaborated ribbon-helix-helix motif involved in DNA recognition. 2012;51(13):2899–2910. doi:10.1002/jcp.22063
92. Schleper C, Kubo K, Zillig W. The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. 1992;89(16):7645–7649. doi:10.1073/pnas.89.16.7645
93. Schmid M, Speiseder T, Dobner T, Gonzalez RA. DNA Virus Replication Compartments. 2014;88(3):1404–1420. doi:10.1128/JVI.02046-13
94. Seed, Kimberly D.; Lazinski, David W.; Calderwood, Stephen B.; Camilli A. A Bacteriophage Encodes Its Own CRISPR/Cas Adaptive Repsonse To Evade Host Innate Immunity. 2013;494(7438):489–491. doi:10.1038/nature11927
95. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awaysz MJ, Chan-Weiher CC-Y, Clausen IG, Curtis BA, De Moors A, et al. The complete genome of the crenarchaeon Sulfolobus solfataricus P2. 2001;98(14):7835–7840. doi:10.1073/pnas.141222098
96. Sinclair RM, Ravantti JJ, Bamford DH. Nucleic and Amino Acid Sequences Support Structure-Based Viral Classification. 2017;91(8):1–13.
97. Stedman KM, Clore A, Combet-Blanc Y. Biogeographical diversity of archaeal viruses. In: Prokaryotic Diversity: Mechanisms and Significance:
98. Stedman KM, DeYoung M, Saha M, Sherman MB, Morais MC. Structural insights into the architecture of the hyperthermophilic fusellovirus SSV1. 2015;474:105–109. doi:10.1016/j.virol.2014.10.014
99. Stedman KM, Schleper C, Rumpf E, Zillig W. Genetic requirements for the function of the archaeal virus SSV1 in Sulfolobus solfataricus: Construction and testing of viral shuttle vectors. 1999;152(4):1397–1405.
100. Stedman KM, She Q, Phan H, Arnold HP, Holz I, Garrett RA, Zillig W. Relationships between fuselloviruses infecting the extremely thermostable archaean Sulfolobus: SSV1 and SSV2. 2003;154(4):295–302. doi:10.1016/S0923-2508(03)00074-3
101. Stetter KO. History of discovery of the first hyperthermophiles. 2006;10(5):357–362. doi:10.1007/s00792-006-0012-7
102. Torarinsson E, Klenk HP, Garrett RA. Divergent transcriptional and translational signals in Archaea. 2005;7(1):47–54. doi:10.1111/j.1462-2920.2004.00674.x
103. Wiedenheft B, Stedman K, Roberto F, Willits D, Gleske A-K, Zoeller L, Snyder J, Douglas T, Young M. Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. 2004;78(4):1954–1961. doi:10.1128/JVI.78.4.1954
104. Wu Z, Liu J, Yang H, Xiang H. DNA replication origins in archaea. 2014;5(APR):1–7. doi:10.3389/fmicb.2014.00179
105. Zhai B, DuPrez K, Doukov TI, Li H, Huang M, Shang G, Ni J, Gu L, Shen Y, Fan L. Structure and Function of a Novel ATPase that Interacts with Holliday Junction Resolvase Hjc and Promotes Branch Migration. 2017;429(7):1009–1029. doi:10.1016/j.jmb.2017.02.016
106. Zhan Z, Zhou J, Huang L. Site-Specific Recombination by SSV2 Integrase: Substrate Requirement and Domain Functions. 2015;89(21):10934–10944. doi:10.1128/JVI.01637-15
107. Zhang J, Kasciukovic T, White MF. The CRISP R Associated Protein Cas4 Is a 5′ to 3′ DNA Exonuclease with an Iron-Sulfur Cluster. 2012;7(10). doi:10.1371/journal.pone.0047232
108. Zhu R, Hao Z, Lou H, Song Y, Zhao J, Chen Y, Zhu J, Chen PR. Structural characterization of the DNA-binding mechanism underlying the copper(II)-sensing MarR transcriptional regulator. 2017;22(5):685–693. doi:10.1007/s00775-017-1442-7
109. Zillig W, Kletzin A, Schleper C, Holz I, Janekovic D, Hain J, Lanzendörfer M, Kristjansson JK. Screening for Sulfolobales, their Plasmids and their Viruses in Icelandic Solfataras. 1993;16(4):609–628. doi:https://doi.org/10.1016/S0723-2020(11)80333-4
110. Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, Scholz I. The Sulfolobus—“Caldariella” group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. 1980;125(3):259–269. doi:10.1007/BF00446886
# Appendix: Primers Used

| Primer Name          | Primer Sequencea                  |
|----------------------|-----------------------------------|
| SSVL Long #10F       | CCCCTTTCGCAAACTCCTTAGTACTTATCGC  |
| SSVL Long #12R       | GATGCATTGGAAAGATGGGC              |
| SSVL Long #11R       | GAACAACTTTAACGCAAAGCAGAACGC      |
| SSVL Long #9R        | TTCTTGCAGCAACAGAGGGCG            |
| SSVL Long #9F        | CAATTCATCGTAATTAGCATGAGG         |
| SSVL Long #8F        | CTCTTGAAACCAGCTTCCAGTAATAACC     |
| SSVL Long #8R        | CGCTAAAATAGTTAGTGATATCTAAACC     |
| SSVL Long #7F        | AAAATTGTACCTATCCATTCTTGC         |
| SSVL Long #7R        | GTGTGCTTAGATACACTATTATCCACC      |
| SSVL Long #6F        | CTGTCATTGGCCAGATTGGG             |
| SSVL Long #5F        | GTGTGGGAGTAAACAGCTATATGCG        |
| SSVL Long #6R        | CTTATCGAACATCTATCTAACC           |
| SSVL Long #5R        | CAGGTGCCTTTCCACTGCC              |
| SSVL AmbReg#1 FP     | CTACGTCAACCTCATTACGATTCC         |
| SSVL AmbReg#1 RP     | GGATGAAAAAGTTACAAGAGATAGAGAACCC  |
| SSVL AmbReg#2 FP_Long| GAATGTTATTATCAGCGTGTCTTCAGTGTC  |
| SSVL AmbReg#2 RP_Long| GTGATTAGGTGGAGGTGTTAGCGCGTGACT   |
| SSVL AmbReg#2 FP_Short| CTCTCACCCATTTATCTACG            |
| SSVL AmbReg#2 RP_Short| CAAAGATGCAAGACTGCTCAAGGC       |
| SSVL VP1/VP3 FP      | GCTTAGTGAACGCCCTATTGACCAAGG      |
| SSVL VP1/VP3 RP      | GATTACATCGCCTAACCAGCCTATTG      |
| SSVL_C82FP           | CAATGAGTGGTACGCC                |
| SSVL_B104FP          | ATGTTGAAACCACAAAATAAGG          |
| SSVL_B298Internal   | GGGGCTTATATTGTGACTACCCAG        |
| SSVL_A814Seq         | GACGAAAATAGTTAAGCCCTTTAT        |
| B205 LiPCR_FP        | GGGAAATAGGAAATGGAGTCAG          |
| B205 LiPCR_RP        | CATCGTCTACCTTATGCTTA            |
| SSV10_C59_FP        | ATGAATATTAGATCATATTATAGGCTTTTAT |
| SSV10_C59_RP        | TCATCCCTTTGTCACCCCTCATG         |
| Sequence Name          | Primers                                      |
|-----------------------|----------------------------------------------|
| SSV10_B205_FP         | ATGAGTAGATTATTAACATATTATAATTATGAAAA         | TAG |
| SSV10_B205_RP         | TCACTGCCTGACTCCATTCC                       |
| SSV10_C127_FP         | ATGGAGTCAGGCAGTGAAG                      |
| SSV10_C127_RP         | CTAACAGCAGATATTATGTTTTTGC                  |
| SSV10_D335(INT)_FP    | ATGGGTAATAAAGTTTTCACATTTTG                |
| SSV10_E51_FP          | ATGGTGCCAGCATATGGACC                    |
| SSV10_E51_RP          | TTATACTACGTCTACAGCTTTACATCCG             |
| SSV10_F199_FP         | GTGATTAGTGAGGATTGTGAC                    |
| SSV10_F199_RP         | TCACCTAATGAAGGTTTTCAGCC                 |
| SSV10_E138_FP         | ATGGCGGAAATATATGCTCAG                   |
| SSV10_E138_RP         | CTAATCCTAAGAGTTCTAATCTTTCTTAG            |
| SSV10_F64_FP          | ATGACAGAAGTATTGAAGAGCG                  |
| SSV10_F64_RP          | TCACGCGAGAGTCAC                        |
| SSV10_D60_RP          | CTATCTCTTTTCTTCTCAATCTATCGATC           |
| SSV10_F73_FP          | TGATCGAAACTGGAAATAGTGAAGATGACA           |
| SSV10_F73_RP          | TCATTTGCGGGGAAAAAACC                    |
| SSV10_E49_FP          | TTGAGTTGTGACTTTTGTCACTTG                |
| SSV10 VP3_LIPCR_FP    | TGAGTAGGCCCCCTTTATAAAGTC                |
| SSV10 VP3_LIPCR_RP    | CATCCCTTTCCGCC                          |
| SSV10_A74 FP          | ATGTATACTAAATGCCAACCGAG               |
| SSV10_A74 RP          | CTATTTATGCTTACAGAATTTTTC               |
| KAN-2 FP-1<sup>b</sup> | ACCTACAAACAAGCTCTCATCAACC         |
| R6KAN-2 RP-1<sup>b</sup> | CTACCCCTGGAACACCTACATCT       |
| Univ_7F<sup>c</sup>   | ATTCAGATTTCTGWATWCAGAA                 |
| Univ_8R<sup>c</sup>   | TCSCCTAAGCAGACTCATC                     |

<sup>a</sup>All sequences in the 5′ → 3′ direction

<sup>b</sup>EZ-Tn5™<R6Kyorikan-2> Transposon specific primers

<sup>c</sup>Amplifies structural gene region of all SSVs