JOURNAL OF BACTERIOLOGY, Apr. 2002, p. 1974–1987 Vol. 184, No. 7
0021-9193/02/$04.00+0 DOI: 10.1128/JB.184.7.1974–1987.2002
Copyright © 2002, American Society for Microbiology. All Rights Reserved.

# Complete Genomic Sequence of SfV, a Serotype-Converting Temperate Bacteriophage of *Shigella flexneri*

Gwen E. Allison,† Dario Angeles,‡ Nai Tran-Dinh, and Naresh K. Verma*

*School of Biochemistry and Molecular Biology, Faculty of Science, The Australian National University, Canberra ACT 0200, Australia*

Received 19 September 2001/Accepted 8 January 2002

**Bacteriophage SfV is a temperate serotype-converting phage of *Shigella flexneri*. SfV encodes the factors involved in type V O-antigen modification, and the serotype conversion and integration-excision modules of the phage have been isolated and characterized. We now report on the complete sequence of the SfV genome (37,074 bp). A total of 53 open reading frames were predicted from the nucleotide sequence, and analysis of the corresponding proteins was used to construct a functional map. The general organization of the genes in the SfV genome is similar to that of bacteriophage λ, and numerous features of the sequence are described. The superinfection immunity system of SfV includes a lambda-like repression system and a P4-like transcription termination mechanism. Sequence analysis also suggests that SfV encodes multiple DNA methylases, and experiments confirmed that *orf-41* encodes a Dam methylase. Studies conducted to determine if the phage-encoded methylase confers host DNA methylation showed that the two *S. flexneri* strains analyzed encode their own Dam methylase. Restriction mapping and sequence analysis revealed that the phage genome has *cos* sites at the termini. The tail assembly and structural genes of SfV show homology to those of phage Mu and Mu-like prophages in the genome of *Escherichia coli* O157:H7 and *Haemophilus influenzae*. Significant homology (30% of the genome in total) between sections of the early, regulatory, and structural regions of the SfV genome and the e14 and KpLE1 prophages in the *E. coli* K-12 genome were noted, suggesting that these three phages have common evolutionary origins.**

Temperate bacteriophages of *Shigella flexneri* play an important role in serotype conversion, and their association with antigenic variation has been known for many years (38, 46). The basic O-antigen of *S. flexneri* is referred to as serotype Y and consists of repeating units of the tetrasaccharide *N*-acetylglucosamine–rhamnose–rhamnose–rhamnose (46), which forms the common polysaccharide backbone characteristic of all *S. flexneri* serotypes except serotype VI (9). There are 13 recognized serotypes that vary through the addition of glucosyl and/or O-acetyl groups to different sugars in the tetrasaccharide unit. Bacteriophages SfV, SfII, and SfX and cryptic prophages SfI and SfIV encode the factors involved in glucosylation of the O-antigen, and lysogenization results in conversion of serotype Y strains to serotypes 5a, 2a, X, 1a, and 4a, respectively (2, 3, 6, 16, 26, 27, 35, 50); bacteriophage Sf6 encodes an acetyltransferase and confers conversion to serotype 3b (10, 49). The genetic organization of the serotype conversion and integration-excision modules is highly conserved among the genomes of the glucosylating phages (reviewed in reference 4), and this organization is also conserved in *Salmonella enterica* serovar Typhimurium serotype-converting phage P22 (48).

Lysogenization by bacteriophage SfV confers type V O-antigen modification, which involves the addition of a glucosyl group to rhamnose II of the tetrasaccharide repeat through an α1,3 linkage. The sequence of the SfV O-antigen modification genes $gtrA_V$, $gtrB_V$, and *gtrV* and flanking regions (5.9 kb in total) has been previously reported (26, 27). Similar to the other glucosylating phages, the serotype conversion genes are located immediately downstream of the *attP* site, which is preceded by the *int* and *xis* genes (26, 27). This phage integrates into the *thrW* gene of the host, and the *int attP* region of SfV has been used in the development of an integrative vector that was used to construct recombinant vaccine strains (17). Downstream of the *gtrV* gene, one incomplete and two complete open reading frames (ORFs) are predicted (27). These ORFs are transcribed in the opposite orientation to the serotype conversion genes, and the protein encoded by *orf-3* shows homology to other phage tail fiber assembly proteins (27). SfV *orf-2* and *orf-3* are very similar to *orf-5* and *orf-4*, respectively, of the cryptic SfI prophage in the chromosome of serotype 1a strain Y53 (3). These two ORFs in Y53 are in the same location and orientation with respect to the type I O-antigen modification genes, suggesting that SfV and the cryptic SfI prophage may also share structural modules (3).

Apart from their role in serotype conversion, very little is known about the molecular characteristics of temperate phages of *S. flexneri*. Angeles et al. (G. E. Allison, D. Angeles, P.-T. Huan, and N. K. Verma, submitted for publication) recently reported on the morphology and restriction map of SfV. Electron microscopy of the phage particle revealed that SfV belongs in the family *Myoviridae*. Restriction mapping revealed that the phage genome has *cos* sites at the termini. A 5.7-kb fragment adjacent to the *cos* site was sequenced and predicted to encode five ORFs (Allison et al., submitted). Sequence and

* Corresponding author. Mailing address: Faculty of Science, School of Biochemistry and Molecular Biology, The Australian National University, Canberra ACT 0200, Australia. Phone: 61 2 6125 2666. Fax: 61 2 6125 0313. E-mail: Naresh.Verma@anu.edu.au.

† Present address: Department of Agricultural, Food, and Nutritional Science, University of Alberta, Edmonton, AB, Canada T6G 2P5.

‡ Present address: Department of Microbiology, The National University of Singapore, Singapore 119074.



Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

functional analyses suggested that this section of the phage genome encodes the DNA packaging and capsid morphogenesis proteins. We now report on the complete sequence of the entire genome of bacteriophage SfV, and the preliminary analysis of these data is presented. Our results suggest that the organization of the SfV genome is typical of the lambdoid family of phages, and a functional map of the phage genome has been constructed with numerous features described in detail.

## MATERIALS AND METHODS

**Strains, phage and media.** Bacteriophage SfV was originally induced from *S. flexneri* EW595/52 (27). Bacteriophage stocks were propagated on *S. flexneri* SFL124 (Δ*aroD*), serotype Y (29), and phage purification and DNA extraction were performed as described for phage λ (43). Luria-Bertani broth and agar (43) were used for routine propagation of both *Escherichia coli* and *S. flexneri*, and cultures were grown in a 37°C incubator or an orbital shaker. When necessary, the medium was supplemented with ampicillin at 100 μg/ml.

**Preparation and sequencing of phage genomic DNA.** Initially, DNA sequence was obtained from restriction fragments of the phage genome cloned into pUC18 and pUC19. When constructing recombinant plasmids, the BRESAClean DNA Purification Kit (Geneworks) was used to gel purify DNA fragments when necessary. Restriction enzymes were used in accordance with the manufacturer's (MBI Fermentas and Amersham Pharmacia) directions, and ligation mixtures were transformed into *E. coli*. *E. coli* JM109 was routinely used for the construction and propagation of recombinant plasmids. Plasmid DNA was routinely prepared by alkaline lysis (43). For sequencing, plasmid DNA was further purified by using polyethylene glycol precipitation (Applied Biosystems), and the M13 Forward and Reverse primers, complementary to the multiple cloning sites of pUC18 and pUC19, were initially used to obtain phage sequence. When necessary, sequence was determined directly from phage genomic DNA, which was prepared as outlined for phage λ and purified by dialysis (43). Primers for primer walking were obtained from Life Technologies. Plasmid and phage DNA sequence was obtained using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, and reactions were conducted in a GeneAmp 2400 thermal cycler in accordance with the manufacturer's (Perkin Elmer) protocol. Reactions were run on an ABI Prism 377 Automated Sequencer at the Biomolecular Resource Facility in the John Curtin School of Medical Research, The Australian National University.

**Sequence assembly and analysis.** DNA sequences were assembled into contigs by using the Genetics Computer Group (GCG, University of Wisconsin) Fragment Assembly System, which is available through the Australian National Genomic Information Service. Assignment of ORFs was conducted with the ORF Finder program, which is accessible through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/gorf/gorf.html); WebGeneMark.HMM (32) (http://genemark.biology.gatech.edu/GeneMark/whmm.cgi); and the GCG Frames program. Additional nucleotide and protein analyses were performed with various GCG programs and other web-based programs as indicated elsewhere in the text.

**Functional analysis of *orf-41*.** *Bam*HI fragment E (Allison et al., submitted) was cloned into the *Bam*HI site of pUC19 to construct pNV724. Plasmid pNV724 was cut with *Sma*I, and the 1.0-kb fragment (nucleotides [nt] 29397 to 30444) was cloned into the *Sma*I site of pUC19 to create pNV910 and pNV911. Plasmids pNV910 and pNV911 were transformed into *Escherichia coli* GM42 (*his dam-3*) (34) to create B1045 and B1046, respectively. The chromosomal DNA from lysogenic and nonlysogenic *S. flexneri* was prepared by the procedure outlined by Bastin et al. (6), digested with restriction enzymes, and subjected to agarose gel electrophoresis.

**Nucleotide accession number.** The nucleotide sequence reported in this paper has been assigned accession number AF339141 in the GenBank database.

## RESULTS AND DISCUSSION

**Genomic sequence of SfV and analysis.** The genome of SfV is 37,074 bp, and the average GC content of the entire genome (50.8 mol% GC) is similar to that of *Shigella* (50 mol%) (7). The DNA sequence was analyzed for the presence of ORFs, and the corresponding proteins were compared with the nonredundant protein databases. A total of 53 ORFs are predicted from the sequence (Table 1 and Fig. 1), and protein-coding regions occupy 92.2% of the genome. Most (ca . 76%) of the genome is predicted to be transcribed to the right; approximately one-quarter of the genome, including the serotype conversion and *attP*, *int*, and *xis* genes, is transcribed in the opposite direction (Fig. 1). Intergenic regions were compared against the nonredundant nucleic acid databases and analyzed for the presence of regulatory sequences. The results of the analysis are discussed below, and the locations and sequences of the predicted Rho-independent terminators are summarized in Table 2 and Fig. 1. The genome was also scanned for the presence of tRNAs with tRNAScan (SE 11, 31; http://www.genetics.wustl.edu/eddy/tRNAscan-SE), but no tRNA genes were identified.

A tentative functional map of the SfV genome was derived from the analyses (Fig. 1). The order of the genes in the SfV genome and the putative transcriptional map and regulatory mechanisms are similar to those of bacteriophage lambda (8). Various features of the SfV genomic sequence and the significance of this homology are described below.

**Phage structural and morphogenesis genes.** The morphology and restriction map of SfV were recently reported (Allison et al., submitted). Electron microscopy of the phage particle revealed an isometric head (ca . 50 nm in diameter) and a long contractile tail (ca . 105 nm in length), characteristic of group A1 morphology in the family *Myoviridae*. SfV is therefore in the same morphology group as phages Mu and P2 (1), as well as serotype-converting phage SfII (35). Restriction mapping and sequence analysis revealed that the phage genome has *cos* sites at the termini. A 5.7-kb fragment adjacent to the *cos* site was sequenced and predicted to contain five ORFs (Allison et al., submitted). Database homology searches suggested that *orf-1*, *orf-2*, and *orf-3* encode the phage small and large terminase subunits and the portal protein, respectively (Table 1). The N-terminal sequence of the capsid protein was determined and corresponded to amino acids (aa) 116 to 125 of the protein encoded by *orf-5*. Functional analysis of *orf-4* indicated that it encodes the phage capsid protease that processes the capsid protein. While a Rho-independent terminator is predicted immediately downstream of *orf-5* (Fig. 1 and Table 2), it is likely that all of the late genes form one transcriptional unit, similar to the situation in phage λ (8).

Analysis of the proteins encoded by *orf-5* through *orf-22* suggests that this region of the genome is involved in the phage tail structure and assembly (Table 1). *orf-10*, *-11*, *-15* to *-20*, and *-22* are homologous to the tail genes of phage Mu and Mu-like prophages in the *Haemophilus influenzae* and *E. coli* O157:H7 genomes (Table 1); *orf-8*, *-9*, and *-13* do not show any significant homology to other proteins in the databases. The homology to phage Mu is consistent with the fact that SfV is in the same morphology group as Mu (Allison et al., submitted). While phage Mu has been studied extensively over the years, relatively little is known about the virion assembly process, in particular, tail structure and assembly. Several earlier reviews were written on this topic (25), and Grimaud (15) has recently summarized the roles of the different genes that are indicated in Table 1.

*orf-19* through *orf-22* encode proteins with homology to those encoded by prophage e14, section 104, of the *E. coli*

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

TABLE 1. Analysis of predicted ORFs and proteins of SfV

| ORF (gene name) | Gene coordinates and orientation | Gene product | | | Function (reference[s]) | Related phage and bacterial proteins | | |
|---|---|---|---|---|---|---|---|---|
| | | Size (aa) | Molecular mass (kDa) | pI[d] | | Protein(s) (size and origin) | GenBank accession no. | BlastP e value (% positives)[c] |
| 1 | 68→562 | 164 | 17.9 | 10.4 | Small terminase subunit | Hypothetical protein Z1853 (118 aa: *E coli* O157:H7 prophage CP-933C) | AE005328 | 1e-10 (59) |
| | | | | | | Hypothetical protein (159 aa; Bacteriophage GMSE-1) | AF311659 | 1e-10 (46) |
| | | | | | | ORF9 (122 aa; *H. influenzae*) | AF198256 | 9e-10 (54) |
| 2 | 559→2292 | 577 | 65.3 | 5.3 | Large terminase subunit | YmfN (455 aa; *E. coli*) | AE000214 | 0.0 (97) |
| | | | | | | Terminase large subunit (563 aa; *Pseudomonas* phage D3) | AF165214 | e-135 (62) |
| | | | | | | Terminase large subunit ECs1598 (553 aa; *E. coli* O157:H7) | AP002555 | 7e-67 (48) |
| 3 | 2392→3666 | 424 | 48.3 | 7.9 | Portal protein | ORF25 (416 aa; *Bacillus* phage phi-105) | AB016282 | 7e-51 (55) |
| | | | | | | Phage phi-105 ORF25-like protein (ORF25 410 aa; *H. influenzae*) | AF198256 | 2e-37 (50) |
| | | | | | | Putative portal protein MIr8522 (410 aa; *Mesorhizobium loti*) | AP003014 | 2e-36 (50) |
| | | | | | | Putative portal protein ECs1592 (403 aa; *E. coli* O157:H7) | AP002555 | 1e-35 (47) |
| 4 | 3659→4261 | 200 | 22.7 | 4.9 | Capsid protease | Putative phage protein STM2236 (172 aa; *S. enterica* serovar Typhimurium LT2) | AE008800 | 7e-75 (90) |
| | | | | | | Putative protease ORF209 (209 aa); *Lactobacillus casei* bacteriophage A2 | AJ251790 | 5e-24 (60) |
| | | | | | | Hypothetical protein Lin2577 (194 aa; *Listeria innocua*) | AL596172 | 5e-21 (54) |
| | | | | | | ORF41 (194 aa; phiPV83 prophage *Staphylococcus aureus*) | AB044554 | 2e-20 (55) |
| 5 | 4272→6501 | 409 | 45.8 | 5.1 | Capsid | Major capsid protein GP36 (392 aa; Bacteriophage phi-C31) | AJ006589 | 1e-14 (37) |
| | | | | | | Hypothetical protein CC2783 (341 aa; *Caulobacter crescentus*) | AE005943 | 6e-08 (38) |
| | | | | | | Phage major capsid protein GP36 (467 aa; *Mesorhizobium loti*) | AP003014 | 3e-06 (37) |
| 6 | 5580→5903 | 107 | 12.4 | 4.2 | Unknown | Hypothetical proteins Z1851 and ECs1594 (98 aa; *E. coli* O157:H7 prophages CP-933C and Sp7)[a] | AE005328, AP0025555 | 2e-06 (48) |
| 7 | 6014→6310 | 98 | 11.4 | 7.9 | Unknown | Hypothetical protein 1752p (111 aa; *Agrobacterium tumefaciens*) | AE008026 | 1e-05 (42) |
| 8 | 6285→6791 | 168 | 19.7 | 12.6 | Unknown | | | |
| 9 | 6788→7348 | 186 | 20.8 | 4.2 | Unknown | | | |
| 10 | 7357→7527 | 56 | 6.4 | 10.1 | Unknown | Gp38 (67 aa; *E. coli* phage Mu) | AF083977 | 3e-06 (59) |
| | | | | | | Putative phage protein YPO1241 (64 aa; *Yersinia pestis*) | AJ414147 | 8e-06 (61) |
| 11 | 7511→9007 | 498 | 53.2 | 5.5 | Tail sheath protein | Putative phage tail sheath protein YPO1242 (502 aa; *Yersinia pestis*) | AJ414147 | e-106 (60) |
| | | | | | | Mu-like tail sheath protein GpL (hypothetical protein HI1511, 487 aa; *H. influenzae* prophage) | U32827 | 2e-80 (55) |
| | | | | | | Tail sheath protein GpL (490 aa; *E. coli* phage Mu) | AF083977 | 3e-72 (51) |
| 12 | 9007→9363 | 118 | | | Unknown | Hypothetical protein YPO1243 (122 aa; *Yersinia pestis*) | AJ414147 | 9e-10 (48) |
| 13 | 9363→9632 | 89 | | | Unknown | | | |
| 14 | 9774→11609 | 611 | 65.3 | 10.0 | Tail protein | Hypothetical protein STY4603 (926 aa; *S. enterica* subsp. *enterica* serovar Typhi) | AL627283 | e-107 (56) |
| | | | | | | OrfG (396 aa; *S. enterica* subsp. *enterica* serovar Typhi) | AF153829 | 2e-52 (67) |
| | | | | | | Tape measure protein (937 aa; *Lactococcus lactis* phage TP901-1) | AF252967 | 5e-18 (39) |
| 15 | 11655→12998 | 447 | 48.5 | 5.6 | Tail/DNA circulation protein | DNA circulation protein (hypothetical protein HI1515; 455 aa; *H. influenzae* Rd prophage) | U32827 | 2e-31 (46) |
| | | | | | | Putative DNA circulation protein ECs4983 (456 aa; *E. coli* O157:H7 prophage Sp16[a] [Mu-like]) | AP002567 | 3e-21 (40) |
| | | | | | | Putative phage protein YPO1246 (468 aa; *Yersinia pestis*) | AJ414147 | 2e-19 (46) |
| | | | | | | DNA circulation protein N (64-kDa virion protein 495 aa; *E. coli* phage Mu) | AF083977 | 1e-16 (39) |
| 16 | 12995→14074 | 359 | 39.2 | 4.9 | Tail protein | 43-kDa tail protein P (379 aa; *E. coli* phage Mu) | AF083977 | 8e-45 (50) |
| | | | | | | Putative tail protein ECs4984 (374 aa; *E. coli* O157:H7 prophage Sp16[a] [Mu-like]) | AP002567 | 4e-35 (44) |
| | | | | | | Putative phage tail protein YPO1247 (351 aa; *Yersinia pestis*) | AJ414147 | 2e-20 (44) |
| 17 | 14074→14622 | 182 | 19.6 | 6.5 | Tail protein (15) | Putative phage baseplate assembly protein YPO1248 (198 aa; *Yersinia pestis*) | AJ414147 | 9e-17 (47) |
| | | | | | | Hypothetical protein ECs4985 (204 aa; *E. coli* O157:H7 prophage Sp16[a] [Mu-like]) | AP002567 | 3e-14 (45) |
| | | | | | | Gp45 (197 aa; *E. coli* phage Mu) | AF083977 | 3e-12 (43) |
| 18 | 14622→15047 | 141 | 16.2 | 5.9 | Tail protein (15) | Putative phage protein GP46 YPO1249 (151 aa; *Yersinia pestis*) | AJ414147 | 2e-13 (58) |
| | | | | | | Gp46 (145 aa; *E. coli* phage Mu) | AF083977 | 7e-12 (50) |
| | | | | | | Mu-like Gp46 protein (hypothetical protein HI1519; 135 aa; *H. influenzae* Rd) | U32827 | 2e-10 (57) |

*Continued on following page*

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

TABLE 1—*Continued*

| ORF (gene name) | Gene coordinates and orientation | Gene product | | | Function (reference[s]) | Related phage and bacterial proteins | | |
|---|---|---|---|---|---|---|---|---|
| | | Size (aa) | Molecular mass (kDa) | pI[d] | | Protein(s) (size and origin) | GenBank accession no. | BlastP e value (% positives)[c] |
| 19 | 15034→16092 | 352 | 38.3 | 4.9 | Tail protein (15) | YmfP (hypothetical protein b1152; 263 aa; *E. coli* prophage e14[b]) | AE000214 | e-143 (96) |
| | | | | | | Mu-like Gp47 protein (hypothetical protein HI1520 355 aa; *H. influenzae* Rd prophage) | U32827 | 5e-40 (46) |
| | | | | | | Hypothetical protein ECs4987 (361 aa; *E. coli* O157:H7 prophage Sp16[a] [Mu-like]) | AP002567 | 1e-35 (43) |
| | | | | | | Gp47 (360 aa; *E. coli* phage Mu) | AF083977 | 3e-34 (44) |
| 20 | 16083→16667 | 194 | 21.6 | 4.9 | Tail protein (15) | YmfQ (hypothetical protein b1153; 194 aa; *E. coli* prophage e14[b]) | AE000214 | e-113 (98) |
| | | | | | | Putative phage protein YPO1251 (115 aa; *Yersinia pestis*) | AJ414147 | 1e-09 (49) |
| | | | | | | Hypothetical protein ECs4988 (186 aa; *E. coli* O157:H7 prophage Sp16[a] [Mu-like]) | AP002567 | 1e-08 (42) |
| | | | | | | Hypothetical protein NMA1825 (188 aa; *Neisseria meningitidis* Z2491) | AJ391256 | 4e-08 (43) |
| | | | | | | Gp48 (180 aa; *E. coli* phage Mu) | AF083977 | 2e-07 (42) |
| 21 | 16671→17321 | 216 | 22.6 | 5.3 | Unknown | Orf5′ (170 aa; *S. flexneri* cryptic prophage SfI) | AF139596 | 5e-92 (95) |
| | | | | | | YcfK (hypothetical protein b1154; 209 aa; *E. coli* prophage e14[b]) | AE000214 | 3e-53 (64) |
| | | | | | | YfdL (Hypothetical protein b2355; 172 aa; *E. coli* prophage KpLE1[b]) | AE000324 | 2e-37 (74) |
| | | | | | | Hypothetical protein Z0314 (236 aa; *E. coli* O157:H7 prophage CP-933H) | AE005203 | 1e-21 (59) |
| 22 | 17230→17733 | 167 | 18.9 | 4.6 | Tail fiber assembly protein | Orf4 (167 aa; *Shigella flexneri* cryptic prophage SfI) | AF139596 | 4e-85 (94) |
| | | | | | | Hypothetical protein P37 (155 aa; phage APSE-1) | AF157835 | 1e-18 (53) |
| | | | | | | Hypothetical protein YcdD (106 aa; *S. enterica* serovar Typhimurium) | M55342 | 2e-12 (74) |
| | | | | | | Putative tail fiber assembly protein U (175 aa; *E. coli* phage Mu) | AF083977 | 3e-12 (75) |
| 23 (*gtrV*) | 19111←17858 | 417 | 47.7 | 10.0 | Serotype-specific glucosyltransferase (4, 26, 27) | GtrX (416 aa; *S. flexneri* bacterio phage SfX) | L05001 | 9e-66 (55) |
| 24 (*gtrB*) | 20031←19108 | 307 | 34.7 | 7.0 | Bactoprenol glucosyl; transferase (4, 26, 27) | $GtrB_{II}$ (309 aa; *S. flexneri* phage SfII) | AF021347 | e-170 (99) |
| | | | | | | $GtrB_{I}$ (306 aa; *S. flexneri* phage SfI) | AF139596 | e-169 (98) |
| | | | | | | $GtrB_{X}$ (305 aa; *S. flexneri* phage SfX) | AF056939 | e-167 (98) |
| | | | | | | $GtrB_{IV}$ (304 aa; *S. flexneri* prophage SfIV) | AF288197 | e-163 (95) |
| | | | | | | Hypothetical protein b2351 (306 aa; *E. coli* prophage KpLE1[b]) | AE000323 | e-162 (96) |
| 25 (*gtrA*) | 20390←20028 | 120 | 13.2 | 10.4 | Flippase? (4, 26, 27) | $GtrA_{I}$ (120 aa; *S. flexneri* phage SfI) | AF139596 | 4e-63 (99) |
| | | | | | | $GtrA_{II}$ (120 aa; *S. flexneri* phage SfII) | AF021347 | 5e-63 (99) |
| | | | | | | $GtrA_{X}$ (120 aa; *S. flexneri* phage SfX) | AF056939 | 1e-58 (94) |
| | | | | | | Hypothetical protein b2350 (120 aa; *E. coli* prophage KpLE1[b]) | AE000323 | 5e-58 (93) |
| 26 (*int*) | 21815←20652 | 387 | 44.8 | 10.5 | Integrase (4, 26, 27) | Int (387 aa; *S. enterica* serovar Typhimurium phage P22) | AF217253 | 0.0 (94) |
| | | | | | | Putative integrase Z0307 (324 aa; *E. coli* O157:H7 prophage CP-933H) | AE005202 | e-177 (98) |
| | | | | | | Int (387 aa; *E. coli* prophage DLP12) | AE000159 | e-168 (84) |
| 27 (*xis*) | 22041→21692 | 116 | 13.0 | 10.2 | Excisionase (4, 26, 27) | Xis, (116 aa; *S. enterica* serovar Typhimurium phage P22) | AF217253 | 1e-51 (89) |
| | | | | | | Xis (115 aa; *E. coli* strain 586) | X16664 | 4e-18 (61) |
| | | | | | | Xis (115 aa; *S. flexneri* phage SfX) | U82084 | 3e-17 (59) |
| 28 | 22347←22042 | 101 | 12.1 | 4.7 | Unknown | Hypothetical protein b2363 (101 aa; *E. coli* prophage KpLE1[b]) | AE000324 | 2e-51 (99) |
| | | | | | | Hypothetical protein ECs2756 (187 aa; *E. coli* O157:H7) | AP002559 | 3e-19 (78) |
| | | | | | | Hypothetical protein ECs1518 (195 aa; *E. coli* O157:H7 prophage CP-933N) | AP002555 | 3e-18 (76) |
| 29 | 22709←22347 | 120 | 13.7 | 4.6 | Unknown | Hypothetical protein b2362 (120 aa; *E. coli* prophage KpLE1[b]) | AE000324 | 2e-66 (97) |
| | | | | | | Hypothetical protein 1942p (226 aa; *Agrobacterium tumefaciens*) | AE008035 | 2e-12 (49) |
| | | | | | | Hypothetical protein CC1451 (250 aa; *Caulobacter crescentus*) | AE005819 | 1e-11 (46) |
| 30 | 23236←22700 | 178 | 20.3 | 5.1 | Unknown | YfdR (hypothetical protein b2361 187 aa; *E. coli* prophage KpLE1[b]) | AE000324 | 1e-98 (99) |
| 31 | 24188←23364 | 274 | 30.5 | 4.8 | Unknown | YfdQ (hypothetical protein b2360; 274 aa; *E. coli* prophage KpLE1[b]) | AE000324 | e-150 (99) |
| | | | | | | Hypothetical protein XF1649 (273 aa; *Xylella fastidiosa*) | AE003991 | 2e-39 (56) |
| 32 | 24616←24254 | 120 | 13.1 | 5.9 | Unknown | Hypothetical protein b2359 (148 aa; *E. coli* prophage KpLE1[b]) | AE000324 | 1e-62 (99) |
| | | | | | | Hypothetical protein XF1650 (124 aa; *Xylella fastidiosa*) | AE003991 | 8e-17 (60) |
| 33 | 25513←25217 | 98 | 11.2 | 4.9 | Unknown | | | |
| 34 (*cI*) | 26360←25686 | 224 | 25.1 | 6.7 | Repressor | Putative repressor protein C2 (hypothetical protein b1145; 224 aa; *E. coli* prophage e14) | AE000214 | e-131 (99) |
| | | | | | | Repressor protein cI (223 aa; *Bordetella* phage BP3p) | AY029185 | 8e-22 (48) |
| | | | | | | Repressor protein C2 (216 aa; *S. enterica* serovar Typhimurium phage P22) | AF217253 | 4e-16 (53) |
| 35 (*cro*) | 26451→26651 | 66 | 7.3 | 10.1 | Repressor | Hypothetical protein b1146 (165 aa; *E. coli* prophage e14)[b] | AE000214 | 9e-52 (94) |

*Continued on following page*

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

TABLE 1—*Continued*

| ORF (gene name) | Gene coordinates and orientation | Gene product | | | Function (reference[s]) | Related phage and bacterial proteins | | |
|---|---|---|---|---|---|---|---|---|
| | | Size (aa) | Molecular mass (kDa) | pI[d] | | Protein(s) (size and origin) | GenBank accession no. | BlastP e value (% positives)[c] |
| 36 | 26695→27246 | 183 | 20.1 | 4.7 | Unknown | YmfL (hypothetical protein b1147; 189 aa; prophage e14)[b] | AE000214 | 8e-93 (93) |
| | | | | | | Orf33 (156 aa; *Pseudomonas aeruginosa* phage phi CTX) | AB008550 | 2e-04 (44) |
| 37 | 27243→28079 | 278 | 30.1 | 9.3 | Immunity region | Hypothetical protein Orf179 (179 aa; *S. flexneri*) | Z23101 | 8e-15 (58) |
| | | | | | | OrfB (118 or 119 aa; *S. flexneri*) | Z23100 | 5e-11 (60) |
| | | | | | | Putative regulator/cI repressor Z0337 and ECs0300 (185 aa; *E. coli* O157:H7 prophages CP-9331 and Sp2 [a] [P4-like]) | AE005204 AP002551 | 1e-09 (48) |
| 38 | 28072→28308 | 78 | 8.9 | 11.4 | Unknown | | | |
| 39 | 28305→29123 | 272 | 29.3 | 8.4 | Replication and origin | Hypothetical protein Z1337 (400 aa; *E. coli* O157:H7 prophage CP-933M) | AE005288 | 4e-28 (49) |
| | | | | | | Hypothetical protein ECs1073 (363 aa; *E. coli* O157:H7) | AP002554 | 4e-28 (49) |
| | | | | | | YfdO (hypothetical protein b2358; 122 aa; *E. coli* prophage KpLE1[b]) | AE000324 | 2e-07 (87) |
| 40 | 29126→29614 | 162 | 18.7 | 10.5 | Unknown | YfdM (hypothetical protein b2357; 164 aa; *E. coli* prophage KpLE [b]) | AE000324 | 1e-86 (96) |
| 41 (*dam*) | 29614→30267 | 217 | 24.5 | 6.1 | DNA adenine methylase (Dam) | YfdM (hypothetical protein b2356; 102 aa; *E. coli* prophage KpLE1[b]) | AE000324 | 4e-49 (100) |
| | | | | | | Putative DNA methyltransferase (hypothetical protein Z3349; 175 aa; *E. coli* O157:H7 prophage CP-933V) | AE005443 | 4e-05 (41) |
| | | | | | | ORF32 (175 aa; *E. coli* phage VT2-Sa) | AP000363 | 5e-05 (41) |
| | | | | | | Gp62 (175 aa; *E. coli* phage HK97) | AF069529 | 5e-05 (41) |
| 42 | 30264→30590 | 108 | 12.1 | 10.5 | Regulation? | LexA repressor (205 aa; *Providencia rettgeri*), other LexA proteins | X70965 | 3e-13 (61) |
| 43 | 30587→30976 | 129 | 14.2 | 10.1 | Crossover junction endodeoxy ribonuclease | Putative crossover junction endodeoxyribonuclease Z3115 and ECs2751 (119 aa; *E. coli* O157:H7 prophages CP-933U and Sp14 [a] [lambda-like]) | AE005421, AP002559 | 1e-16 (58) |
| | | | | | | Putative endonuclease Z2057 and ECs1777 (119 aa; *E. coli* O157:H7 prophages CP-933O and Sp9 [a] [lambda-like]) | AE005344 AP002556 | 3e-16 (54) |
| | | | | | | Putative endonuclease Z6061 and ECs2268 (119 aa; *E. coli* O157:H7 prophages CP-933P and Sp12 [a] [lambda-like]) | AE006460 AP002557 | 1e-15 (56) |
| 44 | 30996→31805 | 269 | 30.2 | 9.6 | Unknown | Phage-related protein XF2294 (242 aa; *Xylella fastidiosa*) | AE004041 | 8e-12 (50) |
| | | | | | | KilA (266 aa; *E. coli* phage PI) | X15639 | 1e-10 (52) |
| | | | | | | Unknown protein HkbK (165 aa; *E. coli* phage HK620) | AF335538 | 2e-05 (46) |
| 45 | 31885→32802 | 305 | 34.5 | 7.4 | Unknown | Putative cytoplasmic protein STM2240 (329 aa; *S. enterica* serovar Typhimurium LT2) | AE008800 | e-163 (92) |
| | | | | | | Hypothetical protein b1560 (363 aa; *E. coli*) | AE000253 | e-110 (68) |
| | | | | | | Hypothetical protein ECs2195 (360 aa; *E. coli* O157:H7 prophage Sp11 [a]) | AP002557 | e-108 (68) |
| | | | | | | Hypothetical protein Z2100 (349 aa; *E. coli* O157:H7 prophage CP-933O) | AE005347 | e-108 (68) |
| 46 (*Q*) | 32816→33568 | 250 | 27.8 | 8.9 | Antitermination protein Q | Putative Q protein (hypothetical protein b1559, 260 aa; *E. coli* prophage Qin) | AE000253 | e-137 (95) |
| | | | | | | Putative Q protein Z1345 and ECs1524 (273 aa; *E. coli* O157:H7 prophage CP-933M and Sp8 [a]) | AE005288 AP002555 | 6e-42 (54) |
| | | | | | | Putative antitermination protein STY1036 (265 aa; *S. enterica* subsp. *enterica* serovar Typhi) | AL627268 | 5e-34 (49) |
| 47 | 33818→34012 | 64 | 7.4 | 5.0 | Unknown | Hypothetical protein Orf2 (65 aa; *E. coli* phage P27) | AJ249351 | 2e-27 (96) |
| | | | | | | Hypothetical protein Z2059 and ECs1779 (106 aa; *E. coli* O157:H7 prophages CP-933O and Sp9 [a]) | AE005344 AP002556 | 4e-27 (97) |
| | | | | | | Hypothetical protein Z2103 and ECs2192 (94/65 aa; *E. coli* O157:H7 prophages CP-933O and Sp1)[a] | AE005347 AP002557 | 3e-26 (96) |
| 48 | 34162→35214 | 350 | 40.3 | 8.6 | Unknown | Putative Dam methylase Z2060 and ECs1780 (352 aa; *E. coli* O157:H7 prophage CP-933O and Sp9)[a] | AE005344 AP002556 | 1e-172 (90) |
| | | | | | | Hypothetical protein Orf3 (312 aa; *E. coli* phage P27) | AJ249351 | e-155 (90) |
| | | | | | | Putative Dam methylase Gp52 (284 aa; *E. coli* phage N15) | AF064539 | 7e-81 (70) |
| 49 (*S*) | 35291→35626 | 111 | 11.9 | 10.1 | Holin | Putative bacteriophage protein STY2045 (113 aa; *S. enterica* subsp. *enterica* serovar Typhi) | AL627272 | 1e-53 (94) |
| | | | | | | Putative inner membrane protein STM2237 (109 aa; *S. enterica* serovar Typhimurium LT2) | AE008800 | 4e-37 (94) |
| | | | | | | Fels-1 prophage protein STM0906 (114 aa; *S. enterica* serovar Typhimurium LT2) | AE008738 | 3e-26 (65) |
| 50 (*R*) | 35630→36106 | 158 | 17.7 | 10.0 | Lysin | Putative endolysin STY2044 (158 aa; *S. enterica* subsp. *enterica* serovar Typhi) | AL627272 | 1e-85 (97) |
| | | | | | | Putative endolysin Z1876 and ECs1622 (158 aa; *E. coli* O157:H7 prophages CP-933X and Sp8)[a] | AE005330 AP002555 | 1e-80 (94) |

*Continued on following page*

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

TABLE 1—*Continued*

| ORF (gene name) | Gene coordinates and orientation | Gene product | | | Function (reference[s]) | Related phage and bacterial proteins | | |
|---|---|---|---|---|---|---|---|---|
| | | Size (aa) | Molecular mass (kDa) | pI[d] | | Protein(s) (size and origin) | GenBank accession no. | BlastP e value (% positives)[c] |
| | | | | | | Lysin (158 aa; *E. coli* phage HK97) | AF069529 | 3e-79 (93) |
| | | | | | | Lysin (158 aa; *E. coli* phage HK022) | AF069308 | 4e-79 (93) |
| 51 (*Rz*) | 36090→36482 | 130 | 14.5 | 9.2 | Lysis | Putative phage protein STY2043 (130 aa; *Salmonella enterica* subsp. *enterica* serovar Typhi) | AL627272 | 2e-53 (87) |
| | | | | | | Gp23 (119 aa; *E. coli* phage Mu) | AF083977 | 3e-06 (46) |
| | | | | | | Putative protein P14 (129 aa; phage APSE-1) | AF157835 | 2e-04 (44) |
| | | | | | | Hypothetical protein ECs4963 (85 aa; *E. coli* O157:H7 prophage Sp16[a] [Mu-like]) | AP002567 | 0.005 (44) |
| 52 (*Rzl*) | 36301→36639 | 112 | 12.4 | 10.6 | Lysis | Putative protein P16 (109 aa; phage APSE-1) | AF157835 | 6e-16 (66) |
| | | | | | | Hypothetical protein ECs4964 (148 aa; *E. coli* O157:H7 prophage Sp16[a] [Mu-like]) | AP002567 | 3e-13 (54) |
| 53 | 36666→37016 | 116 | 13.1 | 10.7 | Unknown | Hypothetical protein Orf7 (116 aa; *Xenorhabdus nematophilus*) | AJ133022 | 1e-37 (70) |
| | | | | | | Hypothetical protein 19 (124 aa; *Bacillus* phage Phi-105) | AB016282 | 2e-12 (47) |

[a] The prophage and prophage-like elements in *E. coli* O157:H7 Sakai, reported by Hayashi et al. (18) are summarized at the following website: http://genome.gen-info.osaka-u.ac.jp/bacteria/o 157/sptable2.html.
[b] The *E. coli* K-12 prophages were originally reported by Blattner et al. (7) and have been recently summarized by Hayashi et al. (18) at the following website: http://genome.gen-info.osaka-u.ac.jp/bacteria/o157/sptable3.html.
[c] The BlastP values and the percent positives (i.e. percent similarity) reported by the Blast program are included.
[d] pI, isoelectric point.

genome (accession number AE000214; 7) and cryptic prophage SfI in *S. flexneri* Y53 (3) (Tables 1 and 3 and Fig. 1). The *orf-19*- and *orf-20*-encoded proteins are also homologous to phage Mu tail proteins, and the *orf-22*-encoded protein is similar to the tail fiber assembly proteins of other phages (Table 1), as noted by Huan et al. (26). Relative to the nucleotide sequence reported by Huan et al. (26, 27), a few corrections have been noted, which has resulted in the following changes: three amino acid changes in the protein encoded by *orf-3* (currently designated *orf-22*), a frameshift mutation in *orf-2* (currently designated *orf-21*) that increases the size of the encoded protein from 112 to 216 aa, and the completion and correction (resulting in three amino acid changes) of the sequence of *orf-1* (currently designated *orf-20*). As a result of these corrections, additional homology between the *orf-21*-encoded protein and the partial protein encoded by *orf-5′* of cryptic prophage SfI in the Y53 chromosome was observed. The SfV *orf-2*-encoded protein and the SfI *orf-5*-encoded protein were previously reported to overlap by only 66 aa (3), whereas the homology of the *orf-21*-encoded protein extends across the entire length of the partial *orf-5*-encoded protein (Fig. 1).

The general organization of the left half of the SfV genome is similar to that of other phages. The genes involved in DNA packaging/capsid morphogenesis and tail structure/assembly are located in separate clusters that are divided by a Rho-independent terminator between *orf-5* and *orf-6*. In general, the head and tail genes are transcribed in the opposite orientation to the serotype conversion genes, and a Rho-independent terminator is predicted between *orf-22* and *gtrV* (Fig. 1 and Table 2).

**Early regulatory region.** Sequence and protein analysis suggests that SfV utilizes a lambda-like repression system. Early regulatory events in the lambda phages involve the cI repressor and Cro proteins (8). The cI repressor binds to operator sequences up- and downstream of the *cI* gene, which prevents transcription of the lytic genes, promotes lysogeny, and stimulates transcription of the *cI* gene (8). The Cro protein is typically small (<80 aa), binds to the operator sequences upstream of the *cI* gene, and prevents its transcription (8). The *cI* and *cro* genes are adjacent to one another in the phage genome but are transcribed in opposite directions.

The *orf-34*-encoded protein is almost identical to the f224/b1145 protein of the e14 prophage in the *E. coli* genome (Tables 1 and 3 and Fig. 1) and also shows similarity to the cI homologs of phages P22 (Table 1), 434, L, H-19B, and lambda (data not shown), indicating that *orf-34* encodes the cI homolog in SfV. A small ORF, encoding a basic protein of 66 aa, is predicted 90 bp upstream of and in the opposite orientation to *orf-34*. Analysis of the *orf-35*-encoded protein with the GCG Helix-Turn-Helix (HTH) program indicates the presence of a putative HTH motif, typical of DNA binding proteins, from amino acid 12 to amino acid 33. In addition to being almost identical to the C-terminal region of e14 protein b1146 (Table 1), the *orf-35*-encoded protein also shows a low level of homology to the Cro protein of bacteriophage D3 (data not shown), indicating that *orf-35* is the *cro* gene of SfV.

The intergenic region between *cro* and *cI* and the region downstream of *cI* usually contain oR and oL, respectively, which are characterized by the presence of three and two regions of dyad symmetry (8). While three distinct regions of dyad symmetry are not obvious in the intergenic region between the SfV *cI* and *cro* genes, three sets of inverted repeats (IR1 [19 nt], IR2 [17 nt], and IR3 [19 nt]; Fig. 2) are evident and may play the role of oR. One region of dyad symmetry was identified in the intergenic region between *cI* and *orf-33* (nt 25656 to 25673). The GCG Terminator Program also identified the latter as a putative Rho-independent terminator (Fig. 1 and Table 2). Putative promoter sequences were detected upstream of *cro* (Fig. 2); however, no obvious promoter sequences were detected for *cI*. Unlike the situation in lambda, a strong ribosomal binding site is predicted upstream of the ATG start codon of the *cro* gene. Further experiments are

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

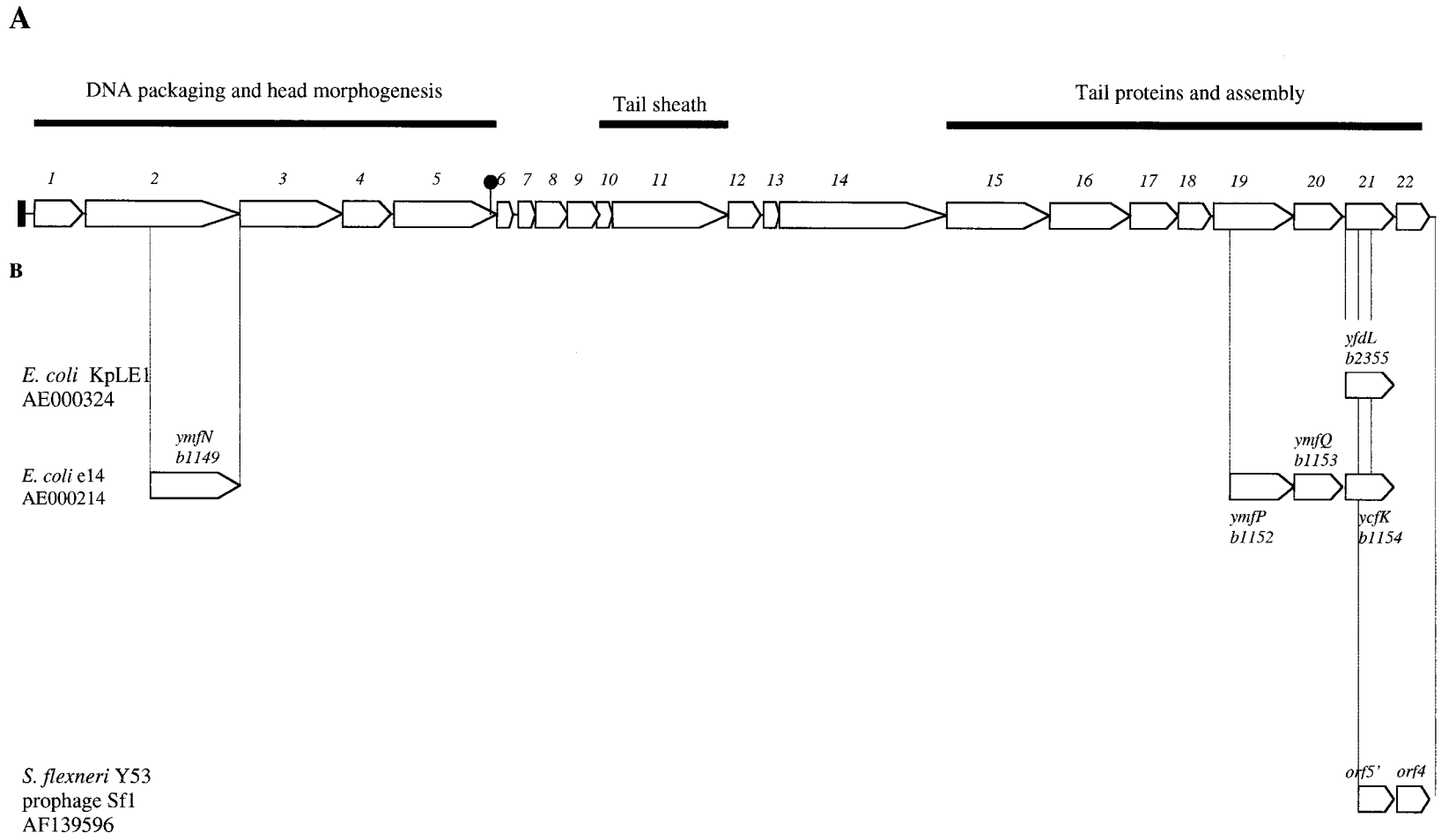


FIG. 1. Genetic map of bacteriophage SfV. (A) Relative locations of the different ORFs. Filled rectangle, *cos* site; filled circles, putative Rho-independent terminators as predicted by the GCG Terminator Program (refer to Table 2 for the sequence). Functional modules are indicated above the genetic map and are based on sequence and experimental analyses. (B) Regions of significant contiguous similarity (>77% identity at the nucleotide level) to other sequences in the data banks.

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

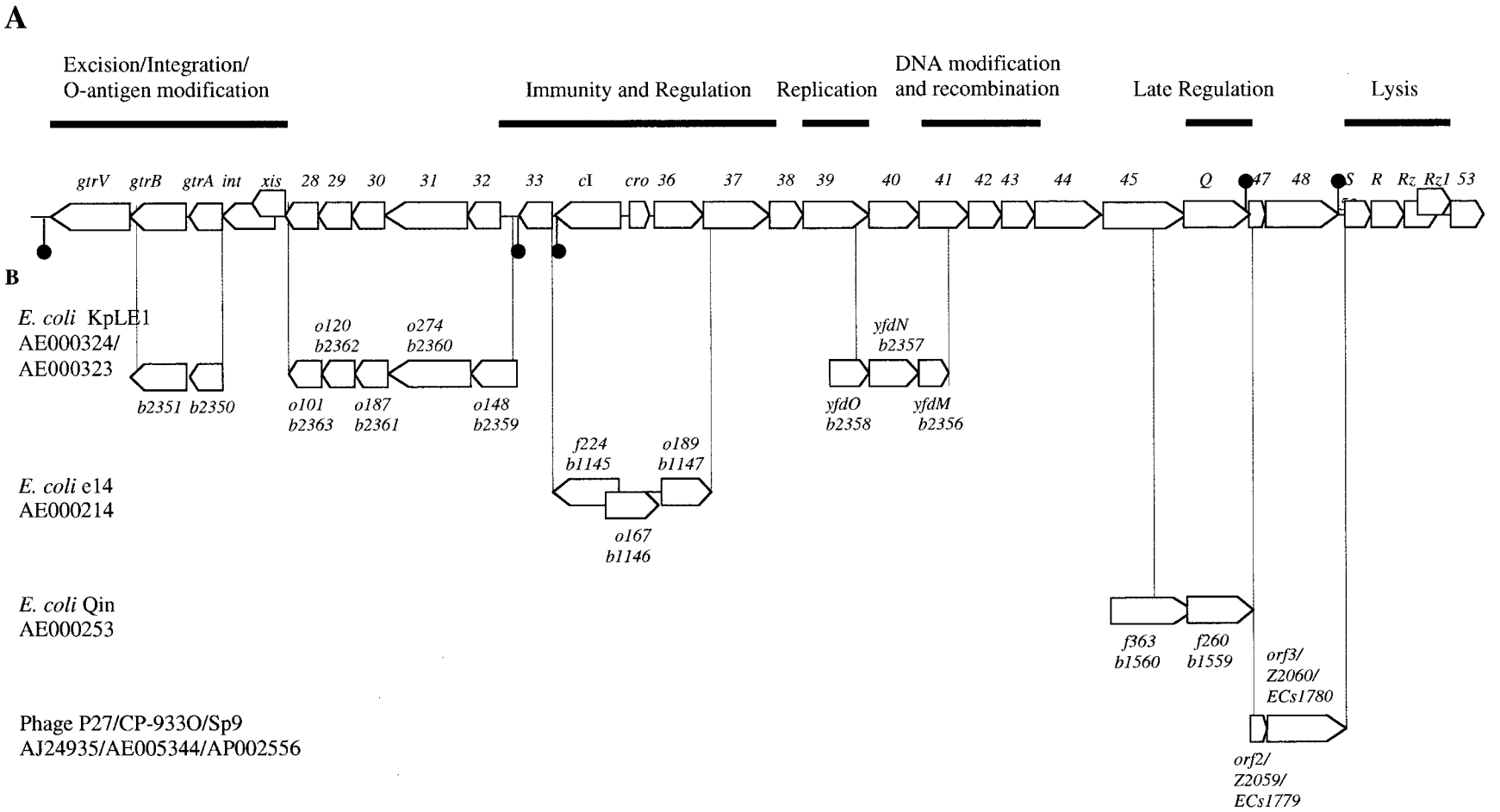


FIG. 1—*Continued.*

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

TABLE 2. Putative Rho-independent terminators[a] in the SfV genome

| Location | Sequence |
|---|---|
| Direct | |
| 5547–5584 | **TGTGCCCGCG**TTCT**GGCGGGCACA**GGAGGTTT TATGCT |
| 33675–33710 | **AACCCGC**CCGCTGA**GCGGGTT**TTTTTGTGCCTT GATG |
| 35222–35253 | **GGCCACG**TTG**CGTGGCC**TTTTTATTTCCAACA |
| Complement | |
| 17807–17770 | **CCCCCCC**GAAACTTA**GGGGGGG**ATCTGCAGT TATAATT |
| 25066–25034 | **ATCAAA**AGTCA**TTTGAT**TTTCCTTTTATGTAT |
| 25673–25649 | **AACCTGC**TTCG**GCAGGTT**TTTTTATACTTGAC |

[a] The Rho-independent terminators were identified by using the GCG Terminator Program. The stem-loop structures are indicated by bolding and underlining, respectively.

required to confirm the role that these features play in the early regulatory events.

Early reports on prophage e14 suggested the presence of a repressor (39, 47), and the sequence analysis presented here suggests that the b1145 protein is the e14 cI repressor homolog. While *orf-35*/*cro* is also predicted by the e14 sequence, the much larger *b1146* ORF has been annotated and overlaps the repressor gene *b1145* (7) (Fig. 1). A *b1146* homolog is also predicted in the SfV sequence. Based on the data presented here and careful analysis of the nucleotide sequence and corresponding proteins in this region, *orf-35*/*cro* has a higher probability of being the coding region.

Additional factors involved in lambda-type regulation, namely, cII, cIII, and N, were not obvious in the protein analyses. The location of *orf-36* and the fact that the corresponding protein is predicted to contain an HTH motif are suggestive of a cII homolog; however, no cII binding sites were identified in the SfV genome. Likewise, a homolog of antitermination protein N was not identified and *nut* sequences were not found. It is expected, however, that antitermination would play a role in transcribing through the Rho-independent terminators predicted in the intergenic region between *cI* (*orf-34*) and *orf-33* and downstream of *orf-33*.

The function of the 2.6-kb region located between *xis* and *orf-33* is unclear. This section of the SfV genome encodes proteins highly homologous to those encoded by section 214 (AE000324; 7) of the *E. coli* genome (Tables 1 and 3 and Fig. 1). The sequence in section 214 shows homology to other bacteriophages (2) and has recently been designated K-12 prophage-like element KpLE1 (18). The proteins encoded by this 2.6-kb fragment were analyzed for the presence of conserved motifs by using the Swiss Institute for Experimental Cancer Research ProfileScan server (http://www.ch.embnet.org/software/PFSCAN_form.html). Weak matches to the RecA and DNA Mismatch Repair 1 motifs were identified in the putative proteins encoded by *orf-30* and *orf-28*, respectively, suggesting that this section of the genome encodes factors involved in recombination. The relative locations of *orf-30* and *orf-28* correspond to those of the recombination genes in other lambda phages (8). Sequence comparisons indicate, however, that another recombination factor is encoded ca. 7 kb downstream, adjacent to the putative origin of replication (refer to the discussion below). The protein encoded by *orf-43* shows homology to putative endonucleases encoded by various prophages in the *E. coli* O157:H7 genome (18, 37) (Table 1). The *orf-43*-encoded protein is also homologous to the RusA proteins encoded by the DLP12 prophage in the *E. coli* genome (GenBank accession number AE000160; BlastP value, 5e-14) and phage 82 (GenBank accession number X92588; BlastP value, 7e-14) (33, 45). RusA is an endonuclease that plays a role in recombination and DNA repair by resolving Holliday junction intermediates (33, 45). RusA homologs have been identified in other phage genomes, where they are typically encoded downstream of the replication-associated genes (45).

**Superinfection immunity in SfV.** Functional and sequence analysis suggests that SfV may have up to three superinfection immunity mechanisms. O-antigen modification confers immunity to SfV (26, 27). Recombinant strains of SFL124 that contain only the O-antigen modification genes $gtrA_V$, $gtrB_V$, and *gtrV* and are completely converted to serotype 5a are immune to further infection by SfV; recombinant strains that contain $gtrA_V$ and *gtrV* or $gtrB_V$ and *gtrV* are only partially converted to serotype 5a (i.e., they display of both serotype Y and 5a O-antigens) and remain sensitive to the phage. Similar SfV immunity and sensitivity phenotypes have been reported for complete and partial conversion, respectively, to serotypes 4a and X (2). O-antigen modification also confers immunity to phages Sf6 (30) and P22 (reviewed in reference 47), both of which use the unmodified O-antigen as the cellular receptor.

In addition to O-antigen modification, sequence analysis suggests that SfV has a typical repressor-mediated lambdoid immunity system (refer to the discussion above). To determine if other superinfection immunity systems exist in SfV, various phage fragments were cloned into pUC18 or pUC19 and introduced into SFL124 (SfV sensitive) and the efficiency of plaque formation on the recombinant strains was determined (G. E. Allison and N. K. Verma, unpublished data). The smallest fragment conferring immunity on SFL124 (efficiency of plaque formation, ca . $10^{-3}$) was a 384-bp *Hin*fI/*Bam*HI fragment (nt 27568 to 27952) from within *orf-37*. Comparison of this sequence against the nonredundant nucleotide database revealed homology to the early region of bacteriophage P4 that mediates superinfection immunity through transcription termination (TT) (Allison and Verma, unpublished). Careful anal-

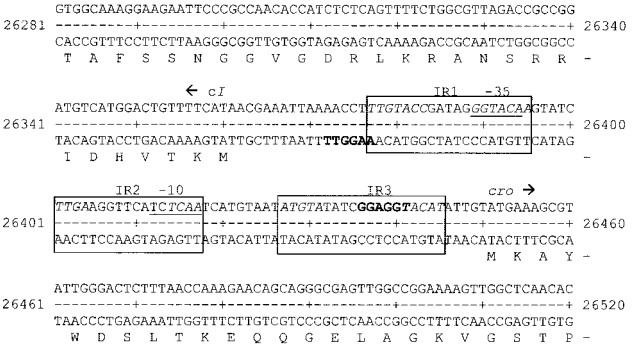


FIG. 2. Intergenic region between the *cI* and *cro* genes of SfV. The putative operator, consisting of three regions of inverted repeats (IR1, IR2, and IR3), is boxed, and the inverted repeats are italicized. Ribosomal binding sites, predicted by WebGeneMark.HMM (32), are in boldface; putative −10 and −35 promoter sequences are underlined.

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

ysis of the *Hin*fI/*Bam*HI fragment indicated that it was predicted to contain the following P4 TT features (Allison and Verma, unpublished): the $P_{LE}$ $\sigma^{70}$ promoter (−35 sequence, TTGATT, nt 27568 to 27573; −10 sequence, TACACT, nt 27591 to 27596); cI RNA containing *seqA*, *seqB*, *seqC′*, and *seqC″* and folding in the conserved secondary RNA structure of the P4 cI RNA; and a nested ORF, *orf-77*, commencing downstream of the cI RNA (the ATG start codon is located at nt 27846) and reading in frame with *orf-37*. In the immune state of phage P4, the cI RNA molecule (69 nt), which is the product of processing of a transcript initiated from constitutive promoter $P_{LE}$, mediates TT and superinfection immunity through RNA-RNA interactions with complementary sequences located up- and downstream in the nascent transcript, thus directly preventing the expression of downstream genes involved in the lytic cycle (14, 42). The cI RNA has a complex predicted secondary structure and contains *seqB*, which is complementary to upstream *seqA* and downstream *seqC′* and *seqC″* (14, 42). $P_{LE}$, *seqA*, *seqB*, and *seqC* are located within the *eta* gene. The *kil* gene is located within and in frame with *eta* and starts downstream of *seqC* (13). Other phages that use a TT-based superinfection immunity system include N15 (40), ϕR73 (41), and phages P1 and P7 (reviewed in reference 19). It is also interesting that *orf-37* homologs are found in *S. flexneri* (12), as well as prophage-encoded proteins in the genome of *E. coli* O157:H7 (Table 1), suggesting that this type of superinfection immunity system may be present in these strains.

**Replication.** The protein encoded by *orf-39* showed homology to hypothetical proteins encoded by various phages in the *E. coli* K-12 and O157:H7 genomes (Table 1). Analysis of the *orf-39*-encoded protein with the GCG HTH program predicted the presence of a putative HTH in the amino terminus (aa 39 to 60). Furthermore, the nucleotide sequence of *orf-39* contains multiple direct repeats. Both characteristics are typical of the replication proteins and origin of replication, respectively, of the lambdoid bacteriophage family (8). It is unknown if other phage proteins are required for replication, but it is possible that *orf-38* and/or *orf-40* are involved.

**Methylases.** Two putative methylases are encoded in the SfV genome. *orf-41* encodes a protein that is homologous to hypothetical proteins in the genomes of *E. coli* (K-12 and O157:H7) and other phages (Table 1), with many of the latter annotated as being similar to DNA methylases. The *orf-41*-encoded protein also showed homology to the previously characterized T1 DNA *N*-6-adenine methylase (28% identity in an 89-amino-acid overlap at the amino terminus) (44). Analysis of the amino acid sequence of *orf-41* revealed that it contains an NPPYSR motif, from amino acid 86 to amino acid 91, that is highly conserved among DNA adenine methylases (Dam) and is involved in binding of the *S*-adenosylmethionine substrate (28).

To determine if the *orf-41*-encoded protein has Dam activity, *orf-41* was cloned into pNV910 and pNV911 on an *Sma*I phage fragment that included 216 and 185 bp up- and downstream, respectively, of *orf-41*. The cloning was initially conducted in JM109 with blue/white selection. Restriction analysis of the corresponding recombinant plasmids from six different transformants revealed that *orf-41* was cloned in the opposite orientation to the vector promoter in all cases. Plasmids pNV910 and pNV911 were subsequently transformed into Dam$^-$ *E. coli* host GM42, resulting in recombinant strains B1045 and B1046. Plasmid DNA extracted from these recombinant strains was digested with *Sau*3AI and *Mbo*I. While both enzymes recognize the same restriction site (↓GATC), *Mbo*I is sensitive to Dam methylation whereas *Sau*3AI is not. Plasmid DNA from control strain B1041 (GM42/pUC18) was restricted by both enzymes, whereas plasmid DNAs from B1045 and B1046 were restricted only by *Sau*3A1 (Fig. 3). The same results were obtained when pNV910 and pNV911 were cloned into Dam$^-$ host strain GM119 (34; data not shown). These data clearly indicate that *orf-41* encodes a DNA adenine methylase. The fact that the gene is expressed when cloned in the opposite orientation to the *lac* promoter in pUC18 suggests that a promoter may be present in the sequence immediately upstream of *orf-41* and/or in the vector. A promoter sequence was noted (−10 signal, TACGGA, from nt 29544 to nt 29549; −35 signal, TTGCGC, from nt 29523 to nt 29528) 63 bp upstream of the ATG start codon. These data also suggest that the SfV Dam methylase may be expressed in lysogens.

The protein encoded by *orf-48* also shows homology to other methylases (Table 1). The proteins encoded by *orf-48* and *orf-47* are very similar to the proteins encoded by phage P27 (nt 33676 to nt 35326, 77% identity in a 1,650-nt overlap) and prophages in the O157:H7 genome (Table 1 and Fig. 1). P27 was recently isolated from a Shiga toxin-producing *E. coli* strain, and *orf-2* and *orf-3* are located upstream of the toxin genes (36). Experiments similar to those conducted on *orf-41* were performed with *orf-48*; however, Dam activity was not detected (data not shown). It is possible that the protein encoded by *orf-47* is involved in nuclease activity, a hypothesis that is strengthened by the observation that the *orf-47* and *orf-48* homologs are found adjacent to one another in phages P27, CP-933O, and Sp9 (Table 1 and Fig. 1). While this gene cassette is conserved among these phages, the location of these genes in the respective genomes is not conserved (18, 36, 37). The significance of the location of *orf-47* and *orf-48* in the SfV genome is discussed below.

To determine if the presence of SfV affects host DNA methylation, we compared the abilities of *Sau*3AI and *Mbo*I to digest the genomic DNA from both cured and lysogenic hosts. EW595/52, which is the lysogenic host used to originally isolate SfV (27), was cured of SfV to create SFL1337 (D. Angeles, G. E. Allison, and N. K. Verma, unpublished data). Southern hybridization, serotype conversion, and phage sensitivity tests indicated that the prophage had been removed from the bacterial chromosome (Angeles et al., unpublished). SFL1, the wild-type parent of serotype Y strain SFL124 (29), was lysogenized by SfV to create SFL1338. SFL1338 converted to serotype 5a and was resistant to SfV (Angeles et al., unpublished). Chromosomal DNAs were extracted from EW595/52, SFL1337, SFL1, and SFL1338 and digested with *Sau*3AI and *Mbo*I. All genomic samples were digested by *Sau*3AI; all samples were resistant to digestion by *Mbo*I (Fig. 3). These data suggest that subtraction or addition of SfV does not affect whether the host DNA is methylated or not and indicate that the *S. flexneri* strains tested encode their own Dam methylase. The importance of Dam methylation in virulence has recently been reported (20). Dam$^-$ mutants of *S. enterica* serovar Typhimurium, as well as Dam overproducers, are avirulent, indicating that the presence and precise amount of Dam are important in

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

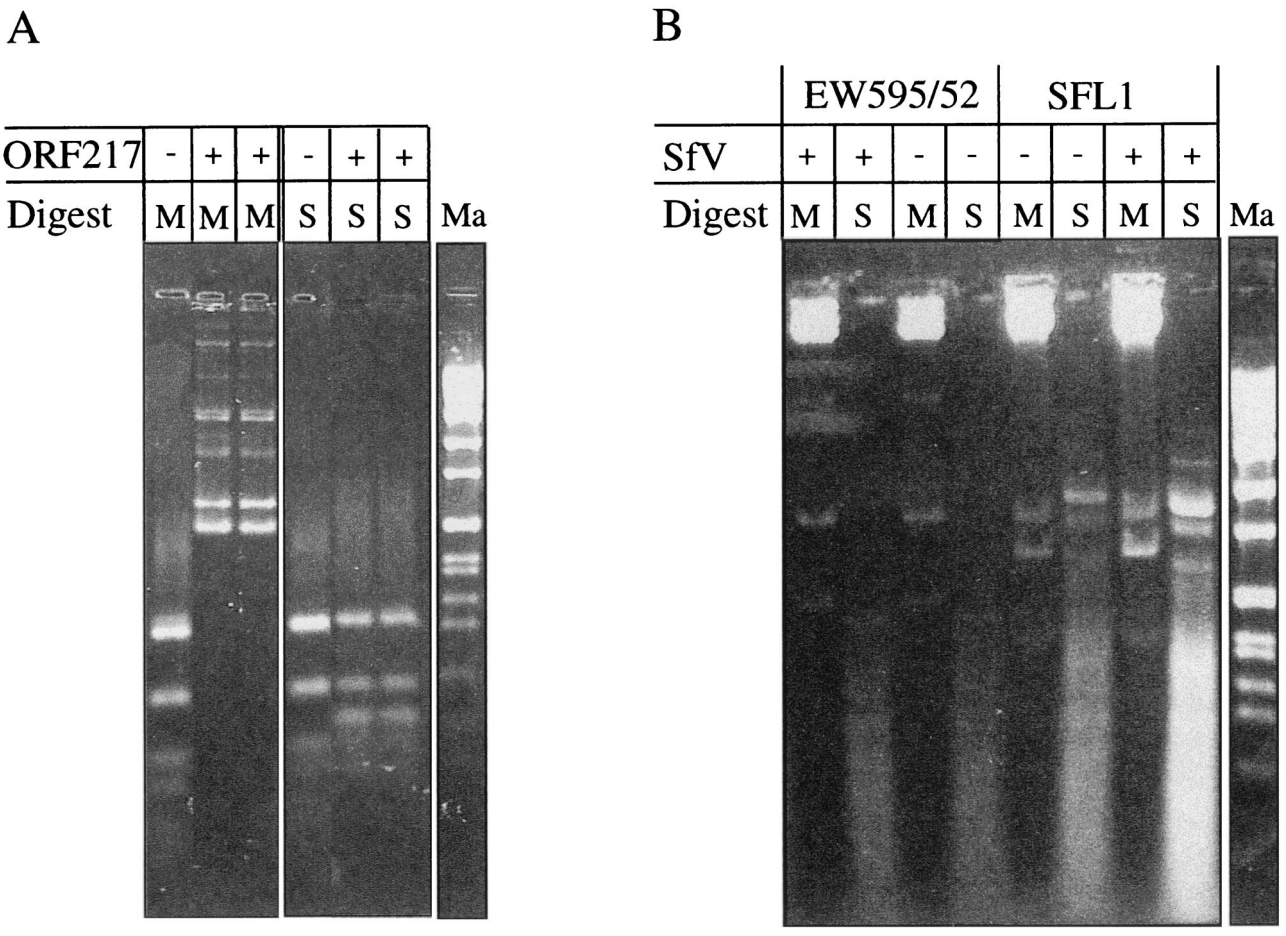


FIG. 3. Functional analysis of *orf-41* (A) and effect of SfV on host DNA methylation (B). The presence and absence of *orf-41* or SfV are indicated by plus and minus signs, respectively. *Mbo*I and *Sau*3AI digests are represented by M and S, respectively. In panel A, plasmid DNAs were extracted from strains B1041 (lanes 1 and 4), B1045 (lanes 2 and 5), and B1046 (lanes 3 and 6). In panel B, total DNA was extracted from strains EW595/52 (lanes 1 and 2), SFL1337 (lanes 3 and 4), SFL1 (lanes 5 and 6), and SFL1338 (lanes 7 and 8). Ma, *Eco*RI-digested SPP-1 molecular weight marker.

the virulence of this organism (20). The data suggest that both EW595/52 and SFL1 encode their own Dam methylase, but it remains to be determined if Dam activity affects *Shigella* virulence and whether the presence or absence of the phage affects the degree to which the bacterial genome is methylated. Dam activity in the host may indicate that acquisition of methylases by SfV plays an important role in propagation of the bacteriophage in the environment.

**Late regulation and lytic genes of SfV.** The late regulatory region of SfV has an organization similar to that of other lambdoid phages. The protein encoded by *orf-46* shares homology with other phage antitermination proteins (Table 1) and has been named Q. A Rho-independent terminator is predicted in the untranslated region downstream of Q (Fig. 1 and Table 2) and is presumably involved in antitermination. *orf-50*, located ca . 2 kb downstream of the *Q* gene, encodes a protein with significant homology to the lysins of HK97, HK022, and putative lysins of prophages in the *E. coli* O157:H7 and *S. enterica* subsp. *enterica* serovar Typhi genomes (Table 1). The protein encoded by *orf-49*, located immediately upstream of *orf-50*, is quite hydrophobic and shows limited homology to the P22, lambda, HK97, and HK022 holin proteins (Table 1 and data not shown). Analysis of the *orf-49*-encoded protein by the TMPred program (23) (http://www.ch.embnet.org/software/TMPred_form.html) predicts the presence of three transmembrane regions. The organization of *orf-49* and *orf-50* and the characteristics of the *orf-49*- and *orf-50*-encoded proteins are consistent with the lytic cassettes of coliphages encoding homologs of the class I holin $S^{\lambda}$ and λ transglycosylase (reviewed in reference 51). *orf-49* and *orf-50* therefore encode the holin and lysin, respectively, of SfV.

Many of these lytic cassettes include the *Rz* and *Rz1* genes (reviewed in reference 51). These two proteins contribute to lysis, but the absolute role they play is unknown (51). The *Rz* gene overlaps or is immediately downstream of the *R* (lysin) gene. The *Rz1* gene, which is usually nested within the *Rz* gene in a +1 reading frame, is a prolipoprotein that is processed at a conserved cysteine residue to yield a small, proline-rich protein. *orf-51* overlaps the lysin-encoding gene and encodes a protein with homology to a hypothetical protein of *S. enterica* subsp. *enterica* serovar Typhi, the GP23 protein of phage Mu, and the P14 protein of phage APSE-1 (Table 1). While the function of these proteins is not known, GP23 and P14 are encoded downstream of the respective phage lysin-encoding

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

TABLE 3. Similarity between the nucleotide sequences of SfV and *E. coli* prophages

| SfV | | | *E. coli* K-12 | | |
|---|---|---|---|---|---|
| Nucleotide positions (total no. bp) | Predicted ORFs[a] | Prophage[c] (accession no. or refs.) | Homologous nucleotide positions | Predicted ORF(s)[a] | % identify at nucleotide level |
| 1152–2237 (1085) | (*orf-2*) | e14 (AE000214) | 7807–8892 | *b1149* | 95 |
| 15305–16998 (1693) | (*orf-19*), *orf-20*, (*orf-21*) | | 9545–11238 | *b1152, b1153*, (*b1154*) | 96 |
| 20514–19074 (1440) | *attP, gtrA, gtrB* | | 7200–8640 | *b1148*[b] b1149[b] | 82 |
| 24838–25022 (184) | Between *orf-32* and *orf-33* | | 4687–4871 | (*b1142*[b]), (*b1143*[b]) | 97 |
| 25573–27230 (1657) | *cl cro, orf-36*, (*orf-37*) | | 5556–7213 | *b1145, b1146, b1147* | 94 |
| 17024–16733 (291) | (*orf-21*) | KpLE1 (AE000324) | 1251–1542 | (*b2355*) | 92 |
| 24645–21981 (2664) | *orf-32, orf-31, orf-30, orf-29, orf-28* | | 2994–5658 | *b2359, b2360, b2361, b2362, b2363* | 96 |
| 29880–29041 (839) | (*orf-39*), *orf-40*, (*orf-41*) | | 1540–2379 | (*b2358*), *b2357*, (*b2356*) | 95 |
| 20515–19090 (1425) | *attP, gtrA, gtrB* | KpLE1 (AE000323) (2,4) | 7201–8626 | *b2350, b2351* | 82 |
| 33679–32464 (1215) | (*orf-45*), *Q* | Qin (AE000253) | 130–1345 | (*b1560*), *b1559* | 88 |

[a] Parentheses indicate that the region of homology starts within an ORF.
[b] e14 ORFs do not correspond to those in SfV.
[c] The *E. coli* K-12 prophages were originally reported by Blattner et al. (7) and have been recently summarized by Hayashi et al. (18) at the following website: http://genome.gen-info.osaka-u.ac.jp/bacteria/o157/sptable3.html.

gene. *orf-52* overlaps *orf-51*, and analysis of the *orf-52*-encoded protein against the Prosite database (http://www.ch.embnet.org/software/PFSCAN_form.html) (5, 24) identified a prokaryotic lipoprotein motif (conserved cysteine residue located at amino acid 19). Numerous proline residues are present in the predicted mature protein (93 aa). While the mature Rz1 proteins are typically 40 aa, larger Rz1 proteins have been reported (51). The organization of *orf-51* and *orf-52* and the characteristics of the *orf-51*-encoded protein suggest that these two genes may be the *Rz* and *Rz1* homologs, respectively, in SfV.

The region between the *Q* gene and the lytic cassette has been identified as a moron insertion site (reviewed in reference 21). Morons are described as gene cassettes that are independently transcribed and typically flanked by transcription initiation and termination signals that would potentially direct expression of the genes even in a repressed prophage (21). Morons typically occur in the late operons of phages and frequently have significantly different nucleotide composition relative to the adjacent genes. While the functions encoded by many morons are unknown, expression of morons in lysogens is proposed to confer a selective advantage on the host (21). Genes encoding Shiga toxins in 933W, VT2-Sa, H-19B, and APSE-1 and a gene encoding a putative DNA adenine methylase (GP52) in N15 have been identified as morons located between the *Q* gene and the lytic cassette in the respective phage genomes. While the function of the *orf-47*-encoded protein homologs is not known, the *orf-48*-encoded protein is homologous to the putative N15 GP52 DNA adenine methylase (Table 1), although no methylase activity was detected (refer to the discussion above). In the SfV genome, putative −10 (TATTGG) and −35 (TTGCTC) sequences were identified 29 and 51 bp upstream, respectively, of the ATG start codon of *orf-47*; a putative Rho-independent terminator is predicted between *orf-48* and *orf-49* (Fig. 1 and Table 2). Analysis of the GC content of *orf-47* and *orf-48* revealed that it is similar to that of SfV and *S. flexneri* (average GC content of 48%); however, that of the region including *orf-48* and *Q* was slightly lower (46% GC content). While the GC content of this region may not be typical, we propose that the general organization and location of *orf-47* and *orf-48* in the SfV genome strongly resemble those of a moron.

**Evolution of serotype-converting bacteriophage SfV.** Analyses of the genome sequence of SfV indicate that the order of the genes in the phage genome and the putative transcriptional map and regulatory mechanisms are similar to those in bacteriophage lambda (8). Interestingly, the proteins involved in the tail structure and assembly are homologous to and organized in a manner similar to those of phage Mu. This observation is consistent with the *Myoviridae* family morphology type reported by Allison et al. (submitted). Regardless of the conserved organization of the genome, the homologies of the specific proteins encoded by SfV suggest a mosaic nature. The mosaicism of phage genomes has been previously reported and has been the topic of two recent reviews (21, 22).

While the SfV genome and corresponding proteins exhibit homology to various phages originating from different morphology groups and various hosts (Table 1; Allison et al., submitted), there is consistent homology between SfV and the e14 and KpLE1 prophages in the *E. coli* K-12 genome (Fig. 1 and Table 3). The segments of homology are largely found in the early and regulatory regions located in the right half of the genome; however, homology to both phages is also observed in the left half of the genome (Fig. 1 and Table 3). It is interesting that contiguous sequences in e14 and KpEL1 are separated into distinct fragments that are positioned at various locations throughout the SfV genome. For example, while *b2356* to *b2360* are contiguous on the KpEL1 prophage, the SfV homologs of *b2359-b2360* and *b2356* to *b2358* occur ca. 5 kb apart on the phage genome (Fig. 1). Furthermore, the e14 fragment corresponding to nt 7807 to 8640 occurs twice in the SfV genome, suggesting that this fragment performs an important

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

function. The amount of SfV DNA that is significantly homologous to these *E. coli* phages is quite substantial (Table 3): ca. 6 kb from e14, 5.2 kb from KpEL1, and 1.2 kb from Qin. In total, approximately 30% of the SfV genome is significantly homologous to e14 and KpEL1, suggesting that these phages have their evolutionary origin in common, and the high degree of homology among the phage fragments suggests recent evolutionary events.

It is of particular interest that the KpEL1 prophage has similarities to other *S. flexneri* serotype-converting phages. The prophage integrase (encoded in section 213) is very similar to the integrase of Sf6 (7). Directly downstream of the KpEL1 *int* gene are serotype conversion genes, $gtrA_{Ec}$, $gtrB_{Ec}$, and $gtrIV_{Ec}$, that have recently been shown to confer partial serotype conversion from Y to 4a on SFL124 (2). Relative to other glucosyltransferase-encoding genes, $gtrIV_{Ec}$ is quite similar to the native *gtrIV* gene of *S. flexneri* (2). These data indicate that this prophage is involved in serotype conversion in *E. coli*. Gene *b2357* is located downstream of $gtrA_{Ec}$, $gtrB_{Ec}$, *and* $gtrIV_{Ec}$; homologs of *b2357* occur in SfV (*orf-40*) and SfII (35). In both SfV and SfII, the *b2357* homolog is located approximately 9 kb upstream of the phage *int* genes, which raises the possibility that SfV and SfII share other modules in addition to those encoding excision-integration and O-antigen modification. The extensive homologies between SfV and putative serotype-converting prophage KpEL1 and the similarity of the O-antigen modification genes in *E. coli* and *S. flexneri* provoke questions regarding the evolution or potential coevolution of O-antigen modification genes and serotype-converting phages in *E. coli* and *S. flexneri*. On this note, it is of interest that the SfV *attP gtrA gtrB* region is also homologous to a region in e14 (Table 3). While the degree of homology at the nucleotide level is similar to that observed for KpEL1 (Table 3), several gaps are introduced, resulting in virtually no similarity between the SfV and e14 proteins encoded in this region (data not shown). It is tempting to speculate, therefore, that e14 was, at one time, involved in serotype conversion.

It has been known for many years that temperate bacteriophages play an important role in the antigenic variation of *S. flexneri* and contribute to its persistence in the environment by providing a means by which to evade the host immune system. Investigation of other serotype-converting phages and their interactions among themselves and with other phages and bacteria will further contribute to our understanding of the environmental and biological characteristics of this human pathogen.

## ACKNOWLEDGMENTS

We thank Peter Reeves for the *E. coli dam* mutants. We also thank the reviewers for valuable suggestions.

This work was supported by the National Health and Medical Research Council of Australia.

## REFERENCES

1. **Ackermann, H.-W.** 1998. Tailed bacteriophages: the order *Caudovirales*. Adv. Virus Res. **51:**135–201.
2. **Adams, M. M., G. E. Allison, and N. K. Verma.** 2001. Characterisation of the type IV O-antigen modification genes in the genome of *Shigella flexneri* NCTC 8296. Microbiology **147:**851–860.
3. **Adhikari, P., G. E. Allison, B. Whittle, and N. K. Verma.** 1999. Serotype 1a O-antigen modification: molecular characterization of the genes involved and their novel organization in the *Shigella flexneri* chromosome. J. Bacteriol. **181:**4711–4718.
4. **Allison, G. E., and N. K. Verma.** 2000. Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. Trends Microbiol. 8.:17–23.
5. **Bairoch, A.** 1992. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **11:**2013–2088.
6. **Bastin, D. A., A. Lord, and N. K. Verma.** 1997. Cloning and analysis of the glucosyltransferase gene encoding type I antigen in *Shigella flexneri*. FEMS Microbiol. Lett. **156:**133–139.
7. **Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1474.
8. **Campbell, A.** 1994. Comparative molecular biology of lambdoid phages. Annu. Rev. Microbiol. **48:**193–222.
9. **Cheah, K.-C., D. W. Beger, and P. A. Manning.** 1991. Molecular cloning and genetic analysis of the *rfb* region from *Shigella flexneri* type 6 in *Escherichia coli* K-12. FEMS Microbiol. Lett. **83:**213–218.
10. **Clark, C. A., J. Beltrame, and P. A. Manning.** 1991. The *oac* gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. Gene **107:**43–52.
11. **Eddy, S. R., and R. Durbin.** 1994. RNA sequence analysis using covariance models. Nucleic Acids Res. **22:**2079–2088.
12. **Faubladier, M., and J.-P. Bouche.** 1994. Division inhibition gene *dicF* of *Escherichia coli* reveals a widespread group of prophage sequences in bacterial genomes. J. Bacteriol. **176:**1150–1156.
13. **Forti, F., S. Polo, K. B. Lane, E. W. Six, G. Sironi, G. Deho, and D. Ghisotti.** 1999. Translation of two nested genes in bacteriophage P4 controls immunity-specific transcription termination. J. Bacteriol. **181:**5225–5233.
14. **Forti, F., P. Sabbattini, G. Sironi, S. Zangrossi, G. Deho, and D. Ghisotti.** 1995. Immunity determinant of phage-plasmid P4 is a short processed RNA. J. Mol. Biol. **249:**869–878.
15. **Grimaud, R.** 1996. Bacteriophage Mu head assembly. Virology **217:**200–210.
16. **Guan, G., D. A. Bastin, and N. K. Verma.** 1999. Functional analysis of the O antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. Microbiology **145:**1263–1273.
17. **Guan, S., and N. K. Verma.** 1998. Serotype conversion of a *Shigella flexneri* candidate vaccine strain via a novel site-specific chromosome-integration system. FEMS Microbiol. Lett. **166:**79–87.
18. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C.-G. Han, E. Ohtsubo, K. Nakayama, et al.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.
19. **Heinrich, J., M. Velleman, and H. Schuster.** 1995. The tripartite immunity system of phages P1 and P7. FEMS Microbiol. Rev. **17:**121–126.
20. **Heithoff, D. M., R. L. Sinsheimer, D. A. Low, and M. J. Mahan.** 1999. An essential role for DNA adenine methylation in bacterial virulence. Science **284:**967–970.
21. **Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens.** 2000. The origins and ongoing evolution of viruses. Trends Microbiol. **8:**504–508.
22. **Hendrix, R. W., M. C. M. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull.** 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA **96:**2192–2197.
23. **Hofmann, K., and W. Stoffel.** 1993. TMbase a database of membrane spanning protein segments. Biol. Chem. Hoppe-Seyler **347:**166–175.
24. **Hofmann, K. P., P. Bucher, L. Falquet, and A. Bairoch.** 1999. The PROSITE database, its status in 1999. Nucleic Acids Res. **27:**215–219.
25. **Howe, M. M.** 1987. Late genes, particle morphogenesis, and DNA packaging, p. 103–157. *In* N. Symonds, A. Toussaint, P. van de Putte, and M. M. Howe (ed.), Phage mu. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Huan, P. T., D. A. Bastin, B. L. Whittle, A. A. Lindberg, and N. K. Verma.** 1997. Molecular characterization of the genes involved in O-antigen modification, attachment, integration and excision in *Shigella flexneri* bacteriophage SfV. Gene **195:**217–227.
27. **Huan, P. T., B. L. Whittle, D. A. Bastin, A. A. Lindberg, and N. K. Verma.** 1997. *Shigella flexneri* type-specific antigen V: cloning, sequencing and characterization of the glucosyltransferase gene of temperate bacteriophage SfV. Gene **195:**207–216.
28. **Kossykh, V. G., S. L. Schlagman, and S. Hattman.** 1993. Conserved sequence motif DPPY in region IV of the phage T4 Dam DNA-[$N^6$-adenine]-methyltransferase is important for *S*-adenosyl-L-methionine binding. Nucleic Acids Res. **21:**4659–4662.
29. **Lindberg, A. A., A. Karnell, B. A. Stocker, S. Katakura, H. Sweiha, and F. P. Reinholt.** 1988. Development of an auxotrophic oral live *Shigella flexneri* vaccine. Vaccine **6:**146–150.
30. **Lindberg, A. A., R. Wollin, P. Gemski, and J. A. Wohlheieter.** 1978. Interaction between bacteriophage Sf6 and *Shigella flexneri*. J. Virol. **27:**38–44.
31. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequences. Nucleic Acids Res. **25:**955–964.
32. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.HMM: new solutions for gene finding. Nucleic Acids Res. **26:**1107–1115.

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest

33. **Mahdi, A. A., G. J. Sharples, T. N. Mandal, and R. G. Lloyd.** 1996. Holliday junction resolvases encoded by homologous *rusA* genes in *Escherichia coli* K-12 and phage 82. J. Mol. Biol. **257:**561–573.
34. **Marinus, M. G., and N. R. Morris.** 1975. Pleiotropic effects of DNA adenine methylation mutation (*dam*-3) in *Escherichia coli* K-12. Mutat. Res. **28:**15–26.
35. **Mavris, M., P. A. Manning, and R. Morona.** 1997. Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri*. Mol. Microbiol. **26:**939–950.
36. **Muniesa, M., J. Recktenwald, M. Bielaszewska, H. Karch, and H. Schmidt.** 2000. Characterization of a Shiga toxin 2e-converting bacteriophage from *Escherichia coli* strain of human origin. Infect. Immun. **68:**4850–4855.
37. **Perna, N. T., G. Plunkett, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, et al.** 2001. Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.
38. **Petrovskaya, V. G., and T. A. Licheva.** 1982. A provisional chromosome map of *Shigella* and the regions related to pathogenicity. Acta Microbiol. Acad. Sci. Hung. **29:**41–53.
39. **Plasterk, R. H., and P. van de Putte.** 1985. The invertible P-DNA segment in the chromosome of *Escherichia coli*. EMBO J. **4:**237–242.
40. **Ravin, N. V., A. N. Svarchevsky, and G. Deho.** 1999. The anti-immunity system of phage-plasmid N15: identification of the antirepressor gene and its control by a small processed RNA. Mol. Microbiol. **34:**980–994.
41. **Sabbattini, P., E. Siz, S. Zangrossi, F. Briani, D. Ghisotti, and G. Deho.** 1996. Immunity specificity determinants in the P4-like retronphage R73. Virology **216:**389–396.
42. **Sabbattini, P. S., F. Forti, D. Ghisotti, and G. Deho.** 1995. Control of transcription termination by an RNA factor in bacteriophage P4 immunity: identification of the target sites. J. Bacteriol. **177:**1425–1434.
43. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, N.Y.
44. **Schneider-Scherzer, E., B. Auer, E. G. de Groot, and M. Schweiger.** 1990. Primary structure of a DNA ($N^6$-adenine)-methyltransferase from *Escherichia coli* virus T1. J. Biol. Chem. **265:**6086–6091.
45. **Sharples, G. J., S. M. Ingleston, and R. G. Lloyd.** 1999. Holliday junction processing in bacteria: insights from the evolutionary conservation of RuvABC, RecG, and RusA. J. Bacteriol. **181:**5543–5550.
46. **Simmons, D. A., and E. Romanowska.** 1987. Structure and biology of *Shigella flexneri* O antigens. J. Med. Microbiol. **23:**289–302.
47. **van de Putte, P., R. Plasterk, and A. Kuijpers.** 1984. A Mu *gin* complementing function and an invertible DNA region in *Escherichia coli* K-12 is situated on the genetic element e14. J. Bacteriol. **158:**517–522.
48. **Vander Byl, C., and A. M. Kropsinski.** 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J. Bacteriol. **182:**6472–6481.
49. **Verma, N. K., J. M. Brandt, D. J. Verma, and A. A. Lindberg.** 1991. Molecular characterization of the *O*-acetyltransferase gene of converting bacteriophage SF6 that adds group antigen 6 to *Shigella flexneri*. Mol. Microbiol. **5:**71–75.
50. **Verma, N. K., D. J. Verma, P. T. Huan, and A. A. Lindberg.** 1993. Cloning and sequencing of the glucosyltransferase-encoding gene from converting bacteriophage X (SFX) of *Shigella flexneri*. Gene **129:**99–101.
51. **Young, R., I.-N. Wang, and W. D. Roof.** 2000. Phages will out: strategies of host cell lysis. Trends Microbiol. **120:**120–128.

Downloaded from http://jb.asm.org/ on September 21, 2019 by guest