The Journal of Microbiology (2010) Vol. 48, No. 2, pp. 213-222
DOI 10.1007/s12275-010-0055-4

Phenotypic Characterization and Genomic Analysis of the *Shigella sonnei* Bacteriophage SP18§

Kyoung-Ho Kim[2†], Ho-Won Chang[1†], Young-Do Nam[1], Seong Woon Roh[1], and Jin-Woo Bae[1*]

[1]*Department of Life and Nanopharmaceutical Sciences and Department of Biology, Kyung Hee University, Seoul 130-701, Republic of Korea*
[2]*Department of Microbiology, Pukyong National University, Pusan 608-737, Republic of Korea*

(Received February 10, 2010 / Accepted April 1, 2010)

A novel bacteriophage that infects *Shigella sonnei* was isolated from the Gap River in Korea, and its phenotypic and genomic characteristics were investigated. The virus, called SP18, showed morphology characteristic of the family *Myoviridae*, and phylogenetic analysis of major capsid gene (gp23) sequences classified it as a T4-like phage. Based on host spectrum analysis, it is lytic to *S. sonnei*, but not to *Shigella flexneri*, *Shigella boydii* or members of the genera *Escherichia* and *Salmonella*. Pyrosequencing of the SP18 bacteriophage genome revealed a 170-kb length sequence. In total, 286 ORFs and 3 tRNA genes were identified, and 259 ORFs showed similarity (BLASTP e-value<0.001) to genes of other bacteriophages. The results from comparative genomic analysis indicated that the enterophage JS98, isolated from human stool, is the closest relative of SP18. Based on phylogenetic analysis of gp23 protein-coding sequences, dot plot comparison and BLASTP analysis of genomes, SP18 and JS98 appear to be closely related to T4-even phages. However, several insertions, deletions, and duplications indicate differences between SP18 and JS98. Comparison of duplicated gp24 genes and the *soc* gene showed that duplication events are responsible for the differentiation and evolution of T4-like bacteriophages.

***Keywords*:** bacteriophage SP18, *S. sonnei* phage, T4-like phage, whole genome analysis, gene duplication

In 1990s, the annual number of cases of shigellosis was estimated to be up to 163.2 million episodes, with 1.1 million deaths in developing countries and 1.5 million episodes in industrialized countries (Kotloff *et al.*, 1999). Although this number has decreased to 90 million episodes and 108,000 deaths per year (World Health Organization, http://www.who.int/vaccine_research/diseases/diarrhoeal/en/index6.html), shigellosis remains one of the most serious endemic diseases in the world.

Shigellosis is caused by bacteria from the genus *Shigella*, and infection is commonly associated with the ingestion of contaminated food and water (Baird-Parker, 1994). Antibiotics are the primary therapy for shigellosis. However, the imprudent use of antibiotics has led to an increase in antibiotic-resistant and multidrug-resistant *Shigella* species in various countries (Bentley *et al.*, 1996; Talukder *et al.*, 2006; von Seidlein *et al.*, 2006; Penatti *et al.*, 2007; Stafford *et al.*, 2007). A surveillance study of six Asian countries showed that antibiotic resistant *Shigella* species are common; the majority of detected *Shigella flexneri* were resistant to amoxicillin and cotrimoxazole, alone or in combination, and 92% of *Shigella sonnei* isolates were resistant to cotrimoxazole (von Seidlein *et al.*, 2006). Multidrug-resistant bacteria pose a significant challenge to clinical treatment, and infection with multidrug-resistant *Shigella* sp. may even result in death (Watson, 2006).

Bacteriophages and their bactericidal activity have been studied in an effort to find a means of controlling bacterial disease (Merril *et al.*, 1996; Hatfull, 2001; Summers, 2001). Especially, as the spread of multidrug-resistant bacteria continues to pose a serious problem, 'phage therapy' is being considered as an alternative method for the treatment of multidrug-resistant bacteria (Sulakvelidze, 2005; Housby and Mann, 2009). Bacteriophages are recognized as the most abundant biological entities on earth, with an estimated number of approximately 10^{31} (Hendrix, 2003). Bacteriophages perform significant roles pertaining to the global carbon and nitrogen cycle in the prokaryotic ecosystem (Weinbauer, 2004). The great diversity of bacteriophages makes their use applicable in fields such as human medicine, as well as in agriculture, veterinary, and food science (Loc Carrillo *et al.*, 2005; McAuliffe, 2007). However, phage therapy presents numerous challenges for application in human pathogenesis because of the limited information concerning the interaction between phage and host (Housby and Mann, 2009). In addition, for the development of effective and curative phages, it is necessary to isolate new phages as of yet uncultivated and to reveal their physiological and genomic characteristics in detail for application of the phages to the treatment of pathogen infection.

S. sonnei, one of the four species of *Shigella* (*S. flexneri*, *S. sonnei*, *S. dysenteriae*, and *S. boydii*), commonly manifests as diarrheal disease in the USA, Canada, and other industrialized countries (Preston and Borczyk, 1994). Seventy-seven percent of cases of shigellosis of industrialized countries were caused by *S. sonnei*, in contrast to 15% in developing countries (Kotloff *et al.*, 1999). Bacteriophages that infect *Shigella* and *S. sonnei* have been utilized mainly for phage typing (Sechter

† These authors contributed equally to this work. * For correspondence.
E-mail: baejw@khu.ac.kr; Tel: +82-2-961-2312; Fax: +82-2-961-0244
§ Supplemental material for this article may be found at http://www.springerlink.com/content/120956

and Cohen, 1990; Bentley *et al.*, 1996) or studied as temperate phages containing genes for Shiga toxin (Beutin *et al.*, 1999; Strauch *et al.*, 2001). Several strains of bacteriophages that infect *S. sonnei* have been isolated and described in previous studies (James *et al.*, 2001; Strauch *et al.*, 2001; Muniesa *et al.*, 2004). Phage analysis based on genomic resolution is one of the most effective and efficient methods for studying bacteriophages because the phage genome dictates the biological characteristics and physiological and physical functions of the phage, including its life cycle and evolution (Merril *et al.*, 2006). Although the 39,044 bp-sized chromosome of temperate bacteriophages Sf6 that infects *S. flexneri* has been published (Casjens *et al.*, 2004) and several other genomes unpublished are present in GenBank, little is known about the genomic structures or whole genome sequences of phages that infect *Shigella*. In this study, we have isolated a T4-like phage, designated SP18, that infects and lyses *S. sonnei*, and investigated its phenotypic characteristics and complete genome sequence in comparison to its phylogenetic neighbors.

Materials and Methods

Phage isolation

S. sonnei KCTC 2518 was obtained from the Korean Collection for Type Cultures (KCTC) and grown in TSB (Difco, USA) at 30°C for 24 h. A water sample was collected from the Gap river, Daejeon, South Korea (36°26´14.5˝N 127°23´39.4˝E). After the removal of particulate with filter paper, the water sample was filtered through a 0.22 μm pore membrane (Millipore, USA). Five microliter of the filtered water sample was added to the *S. sonnei* KCTC 2518 culture, and the mixture was incubated at 30°C overnight to enrich the concentration of bacteriophages. The enriched culture was then filtered through a 0.22 μm pore membrane, and one drop of filtered water was inoculated on a TSB agar plate on which KCTC 2518 had been spread. Plaques formed during incubation at 30°C were transferred to another KCTC 2518 containing plate. A single plaque was transferred at least three times to purify the bacteriophage.

Growth curve

For determination of the growth curve, early exponential phase cultures of strain KCTC 2518 (optical density at 560 nm=0.2) was inoculated by phages at a multiplicity of infection of 0.1. The culture was incubated at 30°C, and the titer of phages and optical density were determined every 10 min.

Electron microscopy

The purified phage particles were fixed on the copper grid with carbon-coated film and negatively stained with 2% uranyl acetate (pH 4.0). The grids were examined with a JEOL JEM-1010 transmission electron microscope operated at 100 kV.

Host range test

Twenty-one representative bacterial strains related to *S. sonnei* KCTC 2518 were used to determine the host range (Table 1). All strains were obtained from the KCTC. For the test, a paper disk containing the bacteriophages in solution was placed on the center of a TSB agar plate (Difco) on which one of the bacterial strains had been spread. After incubation for 24 h at 30°C, the formation of a clear zone around the paper disk was considered positive activity for lysis of the strain by the bacteriophages.

Table 1. Lysis activity of phage SP18

Number	Culture collection number[a]	Species	Lysis by SP18
1	KCTC 2293	*Escherichia coli*	-
2	KCTC 2441^T	*Escherichia coli*	-
3	KCTC 2790	*Escherichia coli*	-
4	DSM 4481^T	*Escherichia blattae*	-
5	DSM 4560^T	*Escherichia hermannii*	-
6	DSM 4564^T	*Escherichia vulneris*	-
7	DSM 13698^T	*Escherichia fergusonii*	-
8	KCTC 2518	*Shigella sonnei*	+
9	KCTC 2009	*Shigella sonnei*	+
10	DSM 5570^T	*Shigella sonnei*	+
11	DSM 4782^T	*Shigella flexneri*	-
12	DSM 7532^T	*Shigella boydii*	-
13	DSM 14848^T	*Salmonella choleraesuis* subsp. *indica*	-
14	DSM 14846	*Salmonella choleraesuis* subsp. *choleraesuis*	-
15	DSM 13772^T	*Salmonella bongori*	-
16	DSM 9386^T	*Salmonella choleraesuis* subsp. *arizonae*	-
17	DSM 9221^T	*Salmonella choleraesuis* subsp. *hountenae*	-
18	DSM 9220^T	*Salmonella choleraesuis* subsp. *salamae*	-
19	ATCC 13311	*Salmonella enterica* subsp. *enterica* serovar *typhimurium*	-
20	ATCC 13076	*Salmonella enterica* subsp. *enterica* serovar *enteritidis*	-
21	DSM 14847^T	*Salmonella choleraesuis* subsp. *diarizonae*	-

Lysis patterns of phage: +, plaques; -, no plaques.
[a] KCTC, the Korean Collection for Type Cultures; DSM, the German Collection of Microorganisms and Cell Cultures; ATCC, the American Type Culture Collection (ATCC); ^{T} type strain.

Extraction of phage genomic DNA

S. sonnei KCTC 2518 was cultured in 1 L of TSB at 30°C to early exponential phase and inoculated with the phage. After lysis was observed, bacterial cells were collected by centrifugation at 7,000 rpm for 30 min at 4°C. The supernatant was filtered through a 0.22 μm-pore filter membrane. The filtrate was incubated for 2 h at 4°C after the addition of NaCl (final concentration, 0.5 M) and polyethylene glycol 8,000 [final concentration, 10% (w/v)]. Phage particles were precipitated by centrifugation at 7,000 rpm for 30 min at 4°C. The supernatant was discarded and the pellet was resuspended in SM buffer (10 mM Tris-HCl; pH 8.0 and 100 mM $MgCl_2$). Phage particles were purified by CsCl-gradient (ρ=1.4, 1.5, and 1.7) centrifugation as previously described (Sambrook and Russell, 2001). After centrifugation, the fraction containing the phages was collected and purified by dialysis, first for 2 h against a buffer containing 1 M NaCl, 50 mM Tris-HCl (pH 8.0), and 100 mM $MgCl_2$, and then against a buffer containing 100 mM NaCl, 50 mM Tris-HCl (pH 8.0), and 100 mM $MgCl_2$ with dialysis tubing overnight. Phage particles were concentrated using Amicon® Ultra-4 centrifugal filter devices (Millipore) at 2,500 rpm for 30 min at 4°C, and then frozen at -80°C for the extraction of genomic DNA. Genomic DNA from SP18 was isolated using a phenol-chloroform extraction method as described previously (Sambrook and Russell, 2001). The DNA concentration was determined in triplicate using a spectrophotometer (Nanodrop Technologies, USA).

Genome sequencing by 454 pyrosequencing

The phage genome was commercially sequenced through 454 pyrosequencing (Macrogen, Korea). Genomic DNA from the phage was sheared and amplified by emulsion PCR before sequencing by synthesis using the massively parallel pyrosequencing protocol (Margulies *et al.*, 2005). 1/16 regions of a PicoTiterPlate device were used for sequencing with the Genome Sequencer FLX (Roche, USA). The average read length was 206 bp and approximately 15,400 reads were assembled *de novo* into a single contig by the Newbler software included with the instrument. 96.4% of sequences were assembled into a single contig and no other contig more than 900 bp in length were assembled. Assembled sequences were examined manually and compared with the result from assembly using the SeqMan program (DNASTAR, USA). After ORF prediction and annotations, breakages of ORFs due to assembly or sequencing error were examined using the ARTEMIS program (Carver *et al.*, 2008).

Annotation

Open reading frames (ORFs) were predicted using Glimmer software version 3.02 (Delcher *et al.*, 1999) based on nucleotide and amino acid sequence alignment searches (BLASTN and BLASTP) compared to the non-redundant NCBI database. Three potential start codons, ATG, TTG or GTG, were used to predict ORFs. Non-coding regions that were predicted to contain no coding sequences or tRNA were extracted and compared to the non-redundant (nr) database in NCBI GenBank with the BLASTX program. Gene assignment and functional classification of ORFs were based on a homology search against the JS98 and T4 genes. The Artemis and ACT programs were used for genome annotation and comparison (Carver *et al.*, 2008). tRNA was predicted with the tRNAscan-SE program version 1.23 (Lowe and Eddy, 1997). The EMBOSS STRETCHER program (Carver and Bleasby, 2003) was used for calculation of ORF similarity between SP18 and JS98.

Analysis of the SP18 genome

Dot plot analysis was performed with 21 phage genomes related to SP18 using the EMBOSS package (Carver and Bleasby, 2003). To compare phage genomes with SP18, the genomes were downloaded from the NCBI phage databases and their protein sequences were used for StandAlone BLAST. For comparison to prokaryotic genomes, 1599 bacterial genomic component sequences (917 chromosomes and 682 plasmids from 846 bacteria or archaea) were downloaded from the NCBI GenBank database in March 2009. BLASTP analysis of both databases was performed using SP18 ORF sequences as a query with an e-value threshold of 0.001.

Phylogenetic analysis

Amino acid sequences of gp23, the major head capsid protein, were collected from related bacteriophages and aligned using the CLUSTAL X program (Thompson *et al.*, 1997). A neighbor-joining tree was constructed with the MEGA 3.1 program (Kumar *et al.*, 2004). Distance matrices were calculated using the JTT scoring matrix and bootstrap analysis was performed 1,000 times.

Nucleotide sequence accession number

The genome sequence of SP18 was deposited in GenBank under the accession number GQ981382.

Results

Isolation and phenotypic characterization of SP18

Phage SP18 was axenically isolated from the Gap River, Daejeon, South Korea as the host by plaque transfer using *S. sonnei* KCTC 2518. During phage isolation, plaques were observed after overnight incubation at 30°C. SP18 produced clear plaques with a diameter of 7-8 mm after 2 days. Infection experiments showed that the latent period of SP18 in rich medium at 30°C was 30 min from inoculation (Supplementary data Fig. 1). The proliferation of phages ceased after approximately 110 min. In order to study its morphology, isolated SP18 was examined by transmission electron microscopy (Supplementary data Fig. 2). The hexagonal head diameter of SP18 was 81×110 nm and the dimensions of the contractile tail (short and long tails in TEM figures) with fibers were 50-110 nm in length and 23-25 nm in width. Based on its morphology, phage SP18 likely belongs to the family *Myoviridae*, whose members typically exhibit an icosahedral head and a contractile tail with fibers. The host-lytic spectrum range of SP18 was determined by plate test with 21 phylogenetically closely related host strains, including *Escherichia*, *Shigella*, and *Salmonella* species (Table 1). The results showed that SP18 had lytic activity only against strains belonging to the species *S. sonnei*.

Phylogenetic analysis of SP18

The major capsid protein, gp23, was used to determine the phylogenetic relationship of SP18 with other phages (Fig. 1). Sequence analysis revealed that gp23 from SP18 had high amino acid (98%) and nucleotide (91%) sequence identities with gp23 from JS10 and JS98, T4-like enterophages that were isolated from human stool (Chibani-Chennoufi *et al.*, 2004b). The phylogenetic tree showed that the three phages form a clade and belong to the pseudo-T-evens rather than the T-evens group, according to the classification proposed by Tetart

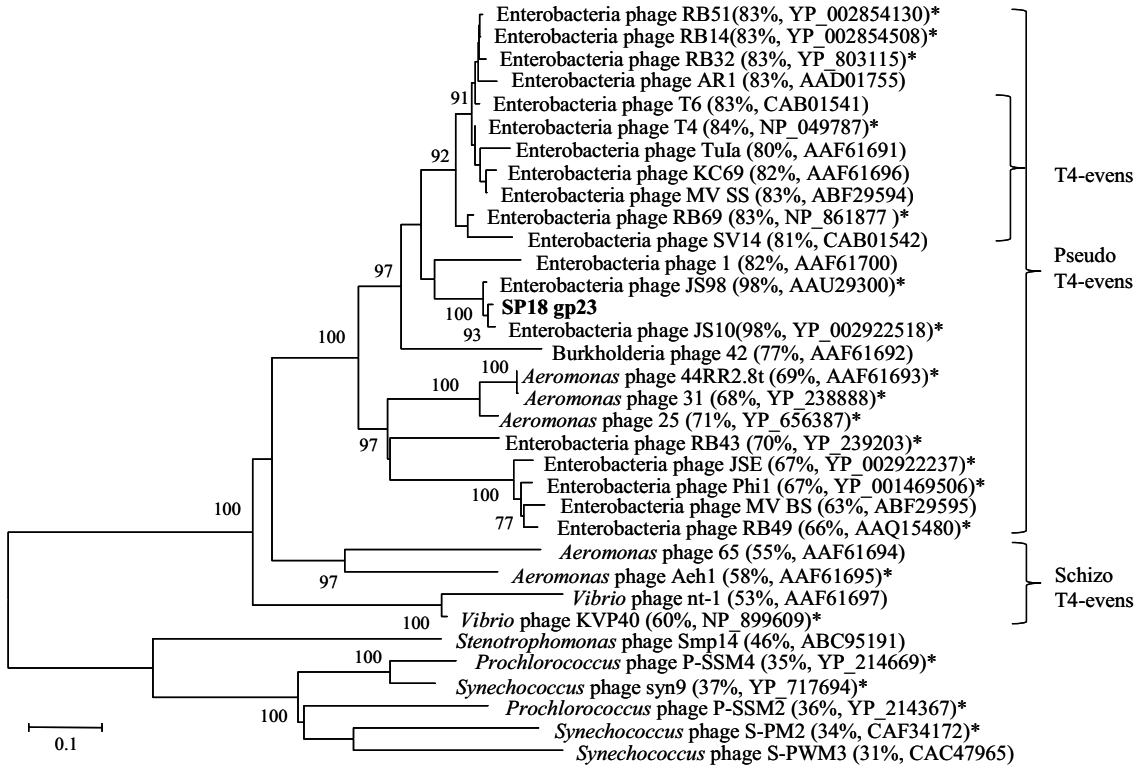

Fig. 1. Phylogenetic relationship of gp23 proteins of *Shigella* phage SP18 and related T4-like phages. The amino acid sequences of the major capsid protein, gp23, were aligned by the CLUSTAL X program. The neighbor-joining tree was constructed based on a distance matrix calculated using the JTT score matrix. Bootstrap values of more than 70% are indicated. Asterisks indicate phage genomes that were completely sequenced. Bar, 0.1 substitutions per site. Accession number of gp23 protein of related phages and % identity with SP18 calculated from BLASTP analysis were indicated.

et al. (2001).

SP18 genome sequence and gene identification

The 170,605 bp long SP18 genome was completely sequenced by 454 pyrosequencing. The coverage of the pyrosequencing reads varied from 7- to 40-fold, but the majorities were more than 15-fold. The final sequence of the SP18 genome assembled into a circle based on terminal redundancy. Similar genome sizes have been reported for the phages JS98 (170,523 bp), JS10 (171,451 bp), T4 (168,903 bp), RB69 (167,560 bp), and RB32 (165,890 bp). The GC content of the SP18 genome is 40.4%, which is lower than that of the *S. sonnei* host (approximately 47%). However, this GC content is higher than T4 (35%) and RB69 (37%), and similar to JS10 (39.5%), JS98 (39.5%), and RB49 (40%).

ORF prediction revealed that SP18 had 286 open reading frames (ORFs) and 3 putative tRNAs that comprised 94.2% of the total genome. The average amino acid length of the ORFs was calculated to be 185 aa. The longest ORF (1,299 aa) was for the gene encoding the proximal subunit of the long tail fiber of gp34. The shortest ORF (37 aa) encoded a hypothetical gene which showed no hit. The number of ORFs in the SP18 genome is higher than those of the other T4-like phages such as T4 (278 ORFs), JS98 (266 ORFs), RB69 (273 ORFs), RB32 (270 ORFs), and RB49 (279 ORFs), but this might be due to the less stringent criteria that were used for annotation. Of the 286 ORFs, 259 showed hits in BLASTP analysis against NCBI nr database with an e-value criterion below 10^{-3}. BLASTP analysis showed that the best hits of the ORFs were related to JS98 (133 ORFs), JS10 (50 ORFs), RB69 (19 ORFs), T4 (16 ORFs), and RB32 (15 ORFs). All hits were phage-related hits except one: the SP18 protein gp195 showed the best hit related to a hypothetical gene of the *Chlorella* virus FR483 (e-value, 1.1e-10; identity, 43%). However, second and third hits were related to putative DNA endonuclease genes of other phages with similar e-values, which implied that the hits were related to genes of bacteriophages. An annotated genome map of SP18 is presented in Fig. 4. We used the *rIIA* gene as the first gene according to the convention used for presentation of T4-type phage genomes.

To examine the relationship between the genes annotated in the SP18 phage genome and bacterial or archaeal genes, BLASTP analysis was performed with a prokaryotic genome database comprised of 846 genomes. Of the 286 ORFs, 47 ORFs showed hits (e-value<0.001) in the prokaryotic genomes database (Supplementary data Table 1). No hits related to genes from the genus *Shigella* were identified as a best hit, but three ORFs showed best hits with *Escherichia coli*, a close relative of *Shigella*. Seven complete genomes belonging to the genus *Shigella* were used for BLASTP comparison and 21 ORFs were found to have significant hits (Supplementary data Table 1).

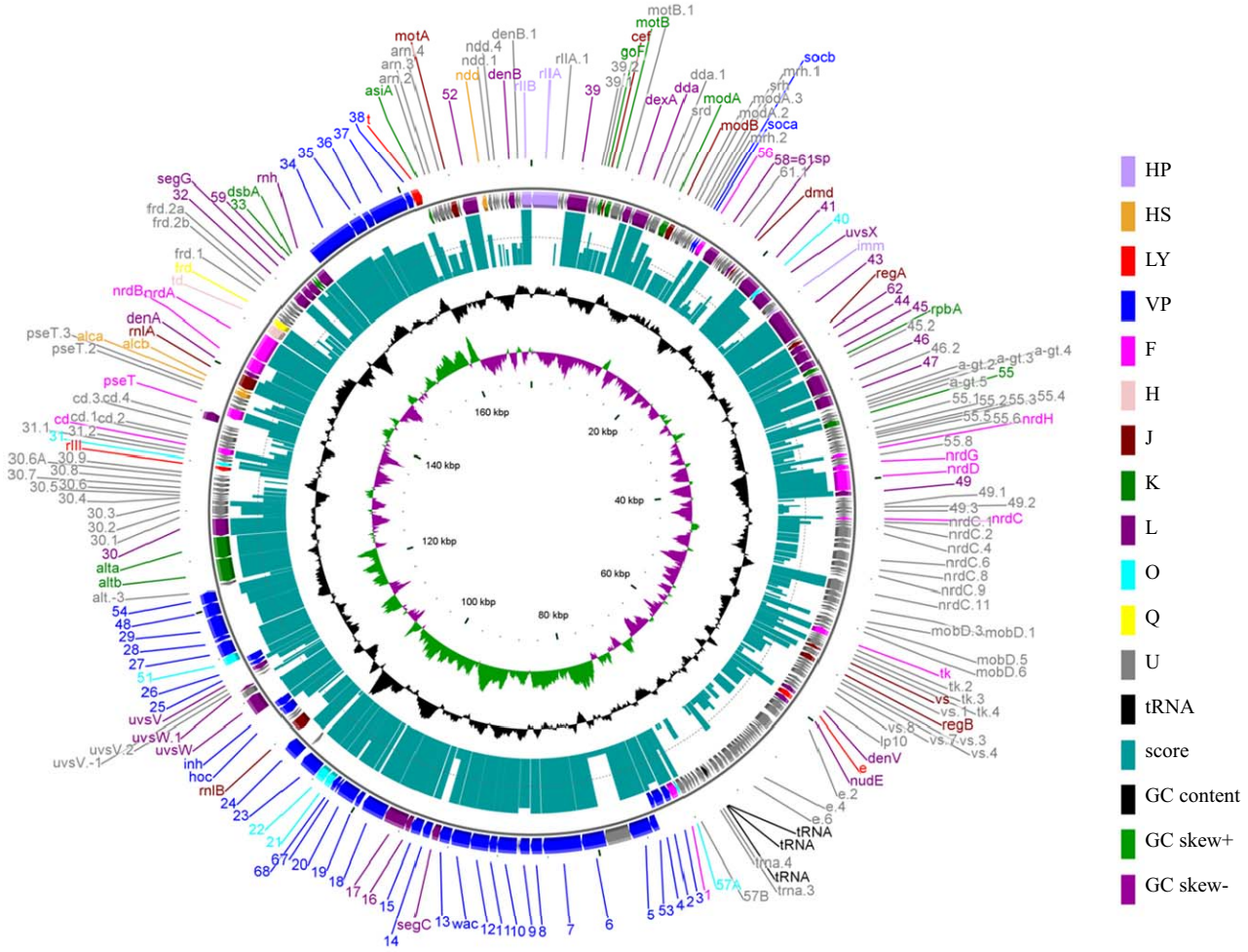

Fig. 2. Circular genome map of SP18. The outer circle represents the ORFs of SP18, and the functions are represented by color. The inner shows the number of hits from BLASTP analysis with 16 T4-like phages. The tallest rectangle indicates the ORF shows similarities with the ORFs of all phages (BLASTP e-value<0.001). The third and fourth circles show the G+C content and GC skew, respectively. The upper and lower graphs denote G+C content that is higher and lower than average, respectively. HP, host or phage interaction; HS, host interaction/shutoff; LY, lysis; VP, viral proteins; F, nucleotide metabolism; H, coenzyme transport and metabolism; J, translation; K, transcription; L, DNA replication, recombination, repair, packaging, and processing; O, chaperonins/assembly catalysts; Q, secondary metabolite biosynthesis, transport and catabolism; U, unknown; tRNA, tRNA; Score, number of hits from BLASTP analysis with 16 T4-like phages.

Comparison with related T4-like phages

The genome structure of SP18 is similar to that of seven related phages: JS98, JS10, RB51, RB32, RB14, T4, and RB69. Dot plot analysis of twenty one related phages including SP18 showed that the eight bacteriophages have a closely related genome structure (Supplementary data Fig. 3). Local BLASTP analysis showed that the ratios of BLASTP hits to the number of ORFs of SP18 were as follows: phage JS98 (82%), JS10 (81%), RB51 (76%), RB32 (74%), RB14 (74%), T4 (72%), RB69 (72%), RB49 (40%), JSE (40%), phi1 (40%), 31 (39%), 44RR2.8 (39%), 25 (38%), RB43 (38%), Aeh1 (33%), and KVP40 (29%). The amino acid (aa) identity of the ORFs of SP18 to the corresponding ORFs of the seven closely related phages was calculated by the STRETCHER program and the average of the values was estimated to be 73% and 72% for JS98 and JS10, respectively, and 55%-57% for all six other phages. 155 ORFs with more than 30% aa sequence identity exist between the eight phages. The genes encoding NrdC, NrdA, gp23, gp21, gp17, gp49, gp55, RegA, and gp56 were highly conserved (mean of average aa identities >80%) between the eight phages. Several other ORFs whose functions are not yet known, such as gp55.4, arn.3, gp39.2, gp30.3, *nrdC.11*, *uvsY.2*, gp57B, and gp30.6, were also shown to be highly conserved between all eight phages (mean of average aa identities >80%).

Comparison of SP18 with JS98

As shown by the results of BLASTP analysis, the genomes of SP18 and JS98 are closely related. Comparison of SP18 and JS98 by local BLASTP analysis showed that 234 ORFs (82%) of SP18 matched those of JS98 with an e-value of 10^{-3} or lower. However, genome comparison also revealed differences between

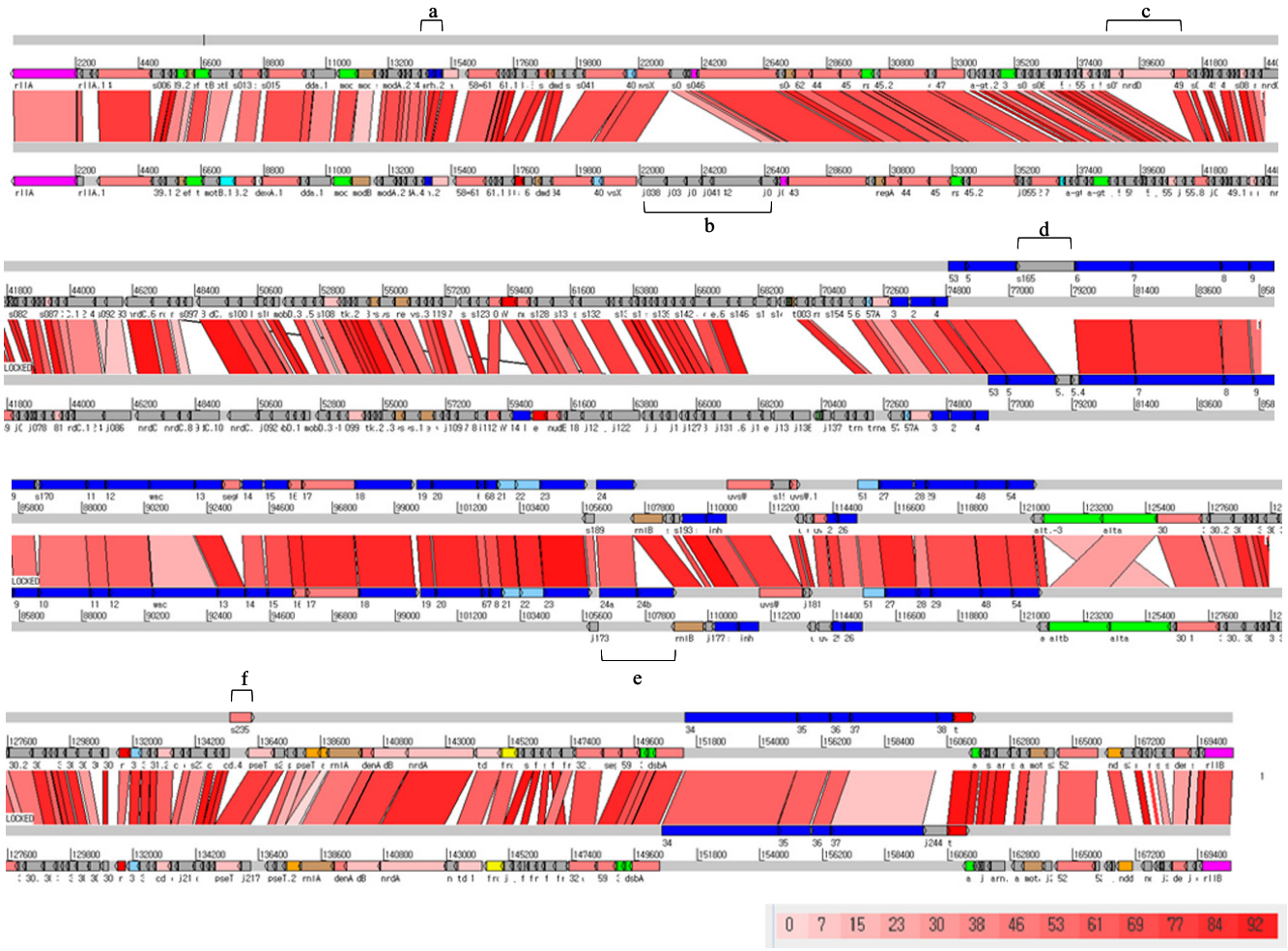

Fig. 3. Genome comparison of SP18 and JS98. Identity calculated by the STRETCHER program is indicated by the intensity of the red color. The assigned function of an ORF is represented by color. Magenta, host or phage interaction; orange, host interaction/shutoff; red, lysis; blue, viral proteins; pink, nucleotide metabolism; pale pink, coenzyme transport and metabolism; brown, translation; green, transcription; light red, DNA replication, recombination, repair, packaging, and processing; light blue, chaperonins/assembly catalysts; yellow, secondary metabolite biosynthesis, transport and catabolism; grey, unknown; pale green, tRNA. (a) Duplication of the *soc* gene in SP18. (b) A deletion in the SP18 genome. (c) An insertion in the SP18 genome. (d) Putative lateral transfer from phage RB43. (e) Duplication of the 24 gene in JS98. (f) Insertion of a putative site-specific intron-like DNA endonuclease in SP18.

two genomes (Fig. 3). Duplication of the *soc* gene encoding small outer capsid protein, was observed in SP18 (Fig. 3a), but not in JS98. Only one copy of the *soc* gene was observed in six related phages, JS98, RB51, RB32, RB14, T4, and RB69, although JS10 was also found to contain two copies of the *soc* gene. A large deletion was observed between uvsX and SP18gp045 in SP18, while this region contained approximately 4,300 bp in JS98 (Fig. 3b). Inserted sequences of various lengths (5,800 bp in RB69, 1,700 bp in RB32, 2,800 bp in T4) have been reported between *uvsX* (RecA-like recombinase gene) and the 43 gene (DNA polymerase gene) in T4-like phages (Chibani-Chennoufi *et al.*, 2004a). The *nrdG* and *nrdD* genes that encode anaerobic NTP reductase subunits were present in SP18 but not in JS98 (or JS10, Fig. 3c). The *nrdD* gene is highly conserved among the T4-like phages and encodes anaerobic ribonucleotide reductase which plays an important role in anaerobic production of the T4 phage (Young *et al.*, 1994). JS98 and JS10 are the only T4-like phages whose genomes have no *nrdD* gene. The *nrdG* gene is also highly conserved, and only JS98, JS10, and the *Aeromonas* phage 31, 44RR2.8t and 25 have no *nrdG* gene among the T4-like phages whose genomes were sequenced completely. The absence of *nrdD* and *nrdG* genes in the JS98 is interesting because these genes also showed similarities with the orthologous genes of bacteria (Miller *et al.*, 2003b). The 687 aa gp165 ORF showed a very low e-value (6e-158) with the hypothetical protein ORF205w of the enterobacteria phage RB43 (Fig. 3d). The gp165 ORF did not show hits with any protein sequences except that of RB43, indicating that this putative gene might have been laterally transferred from phages related to RB43. Duplication of the gp24 gene in JS98, as reported by Zuber *et al.* (2007) was not observed in SP18 (Fig. 3e). SP18 contained genes for the site-specific intron-like DNA endonuclease genes, *segC* and *segG*, but not for the *segA*,

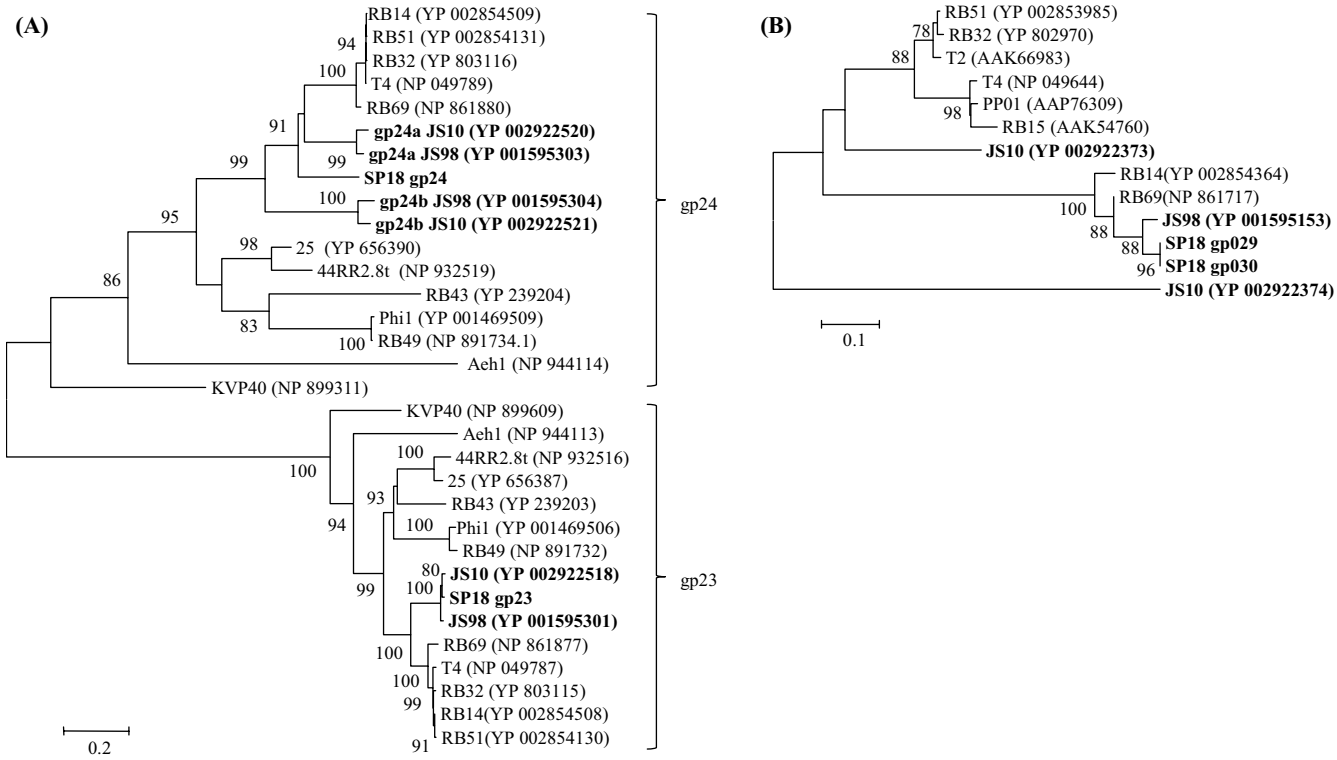

Fig. 4. (A) Comparison of the Gp23 and Gp24 protein sequences of T4-like phages. Duplication of gp24 in JS98 facilitated the diversification of gp24b and gp24a compared to gp23. (B) Comparison of the Soc protein sequences of T4-like phages. SP18 and JS10 both have two copies of the *soc* gene, but their phylogenetic positions are quite different. Phylogenetic trees were constructed using the neighbor-joining method based on the distance matrix calculated using the JTT scoring matrix. Bootstrap analysis was performed 1,000 times and bootstrap values of more than 70% are indicated.

segB, *segD*, *segE* or *segF* genes of T4. SP18 also contained another site-specific intron-like DNA endonuclease gene, gp235, next to the *pseT* gene in the opposite direction (Fig. 3f), while JS98 had no site-specific intron-like DNA endonuclease genes.

Phylogenetic analysis of the duplicated genes

Unlike the other T4-like phages, JS98 and JS10 have two copies of the gp24 gene, which encodes the precursor of the head vertex subunit. The gp24 gene itself was derived from duplication of the gp23 gene, which encodes the precursor of the major head subunit (Haynes and Eiserling, 1996). We examined the phylogenetic relationship of the amino acid sequences of the gp23 and gp24 proteins between T4-like phages (Fig. 4A). In the phylogenetic tree, the degree of sequence divergence of gp24 proteins was higher than that of gp23, which suggests that gp23 proteins were regulated by more stringent constraints than gp24 proteins. It is known that the activity of gp24 can be bypassed by mutations of the gp23 gene (Haynes and Eiserling, 1996). The gp24 gene have evolved more freely than gp23 gene under less strict barriers after duplication. In the case of JS98 and JS10, the duplication event occurred one additional time, to yield the gp24a and gp24b genes. Duplication of protein 24 in JS98 and JS10 is an interesting example of the evolution of T4-like phages. Based on the observation of duplication of gp24, gene duplication and subsequent diversification was suggested to be the mode of evolution for T4-like phages (Chibani-Chennoufi *et al.*, 2004a; Zuber *et al.*, 2007). Because gp24 in SP18 is closely related to gp24 in JS98 and JS10, we were able to investigate the diversification of these phages in detail (Fig. 4A). As the result of duplication, the gp24a and gp24b genes have evolved separately from other genes in the phage genomes. When the gp24 genes of JS98 were compared to the gp24 gene of SP18, the gp24a gene of JS98 showed a closer relationship with the gp24 gene of SP18 than the JS98 gp24b gene. According to the phylogenetic tree, the gp24 gene duplication event appears to have occurred before the differentiation of SP18 and JS98, and the gp24b gene was deleted during the evolution of the SP18. Another explanation is that the gp24 genes of JS98 and JS10 evolved soon after duplication. Although the gp23 genes of SP18 and JS98 form a closely related monophyletic group, the gp24 genes do not form a clear monophyletic group or show the close relationship observed in the gp23 genes. Considering that the relationship of gp23 genes of other T4-like phages was quite similar to that of gp24 genes, the gene duplication event might have had a positive effect on the diversification of the sequences of gp24a and gp24b genes. The existence of two related genes (gp24a and gp24b) may have accelerated the change of both gp24 derivative genes in JS98.

We observed another interesting example of gene duplication in SP18 and JS10. Two copies of the *soc* gene were observed in SP18 with a 92.7% aa identity to each other.

Phage JS10, the closest relative of JS98, also has two *soc* genes. However, the amino acid sequence identity between the two *soc* genes in JS10 is low (41%). Phylogenetic tree analysis indicated that the duplication of *soc* in SP18 might have occurred very recently, while the *soc* genes of JS10 occupied unexpected positions in the tree (Fig. 4B). Because SP18 and JS10 are closely related to each other, the *soc* genes were also expected to be closely related. However, the position of the *soc* gene of JS10 in the phylogenetic tree was unexpectedly different from its position of that of SP18 and JS98, the most closely related phages of JS10 (Fig. 4B). The *soc* gene encodes the small outer capsid gene, which has a length of 76 to 85 aa. Mutation of the *soc* gene caused unstable capsids in the T4 phage and was considered to be an auxiliary gene (Miller *et al.*, 2003b). The *soc* gene was conserved only in the eight phages. Because the *soc* gene is not essential, its evolution was less strictly regulated. Thus, the *soc* genes of the phages create a different tree topology compared to the tree based on gp23 and gp24 genes. In the case of JS10, duplication of the *soc* gene seems to have significantly accelerated the rate of evolution, as shown in the tree in Fig. 4B.

Discussion

Characteristics of SP18

We isolated the phage SP18 from an area of the Gap river that ran through the middle of a city. Because the host, *S. sonnei*, is transmitted to humans by fecal-oral contact, the isolation of phage SP18 specific to *S. sonnei* indicates fecal contamination of the river. Phage JS98, the closest relative of SP18, was first isolated from the stool of a diarrhea patient (Chibani-Chennoufi *et al.*, 2004b). The narrow host range of SP18 has both advantages and disadvantages in its application for phage therapy. Phages specific to only a few strains are not effective for the control of a broad range of pathogenic bacteria. However, this also presents an advantage in that their use does not disrupt the normal flora. In order for the therapeutic application of host-specific bacteriophages to be effective, precise identification of the relevant pathogens before treatment is necessary. Alternatively, a combination of various phages (a phage cocktail) can be used to target a broad range of pathogens. Chibani-Chennoufi *et al*. (2004c) used a phage cocktail composed of four T4-like coliphages isolated from diarrhea patients and environmental water samples and showed that orally applied phage cocktail lysed pathogenic *E. coli* in mice.

Pyrosequencing of the phage genome

454 pyrosequencing, a next generation DNA sequencing technology, can reduce sequencing costs dramatically (Margulies *et al.*, 2005). The technology has been applied to numerous studies to sequence the human genome (Wheeler *et al.*, 2008), as well as to reveal microbial diversity (Roh *et al.*, 2009) and sequence microbial genomes (Herlemann *et al.*, 2009). This high throughput parallel pyrosequencing technique can also be applicable for sequencing of phage genomes. Some phages have genomes of significant size (for example, more than 100 kb) and pyrosequencing offers alternative sequencing tools that are much less expensive than conventional Sanger methods. Above all, conventional cloning has a critical limitation in sequencing phage genomes in that a large number of phage genes are lethal to *E. coli*, which is commonly used in these techniques. Thus, some regions of the phage genomes cannot be fully sequenced by shotgun cloning methods and must be completed by non-cloning methods such as direct PCR, genome walking, and reverse PCR. For example, in the case of JS98, the complete genome sequence could not be obtained by shotgun cloning (Chibani-Chennoufi *et al.*, 2004a) and was finally finished using 454 pyrosequencing (Zuber *et al.*, 2007). In the current study, we used a pyrosequencing approach to sequence the whole genome of phage SP18. A region of only 1/16 of the PicoTiterPlate was sufficient to produce enough reads for assembly of the complete genome with an average coverage of 15-fold.

Comparative genomic analysis of the phage genome

Comparative studies of bacteriophage genome sequences have provided significant information revealing the genetic structure, dynamics of phage population, functions of phage genes, and mechanisms of evolution between phages and their hosts (Hendrix, 2003). More than 520 bacteriophages have been completely sequenced so far (NCBI). The number of completely sequenced phage genomes is much smaller than that of bacteria considering their small genome size, tremendous abundance, and diversity. Even in the case of the most widely studied group of T4-type bacteriophages, the genomes of less than twenty members have been sequenced so far. Although several genomic studies of T4-like phages have been performed (Desplats and Krisch, 2003; Miller *et al.*, 2003a; Chibani-Chennoufi *et al.*, 2004a; Nolan *et al.*, 2006), they are not sufficient to elucidate the evolution and genetic structure of T4-like phages because the number of characterized bacteriophages is too low. The discovery and genomic analysis of novel T4-like phages would provide important and detailed information.

BLASTP analysis of ORFs in the genome of each phage showed that the hit ratios were reduced from 70% to 40% when SP18 was compared with the group, more distant than the eight close phages. This trend coincides with the phylogenetic tree based on gp23 aa sequences, which showed that the eight phages formed a monophyletic group separate from other pseudo T-evens groups. Dot plot analysis also supported the finding that the eight phages were clearly separate from other phages.

The evolution of the T4-like phage genome

It has been proposed that T4-like phages likely evolve by the accumulation of mutations and incidental gene duplication rather than by modular exchanges (Chibani-Chennoufi *et al.*, 2004a). In this study, all significant hits showed the greatest similarity to the genes of bacteriophages, not cellular organisms. Forty-seven ORFs showed significant hits (e-value$<$0.001), but no best hits were with prokaryotic genes, and their similarities or bit scores were much lower than with genes of bacteriophages (Supplementary data Table 1). This implies that lateral gene transfer from bacteria or archaea to SP18 has not occurred. T4-like phages have been reported to be less susceptible to lateral gene transfer (Comeau *et al.*, 2007) than other phages, such as lambda-like phages or those of the *Siphoviridae* group (Pedulla *et al.*, 2003). The complexity of

T4-like phages is proposed to be the reason for restricted lateral gene transfer (Comeau *et al.*, 2007). In contrast, phage-to-phage lateral transfer was observed. For example, two relatively distant bacteriophages, SP18 and RB43, were found to have conserved hypothetical genes (e-value, 5.9e-158; identity, 44.2%).

In this work, we have detailed the physiological and genomic characteristics of phage SP18. This phage efficiently lyses *S. sonnei* and has a morphology characteristic of T4-like phages. Whole genome sequencing revealed a 170 kb-long genome. Genomic analysis showed that SP18 is closely related to JS98, but also shows differences in the genome, such as instances of deletion, insertion, and duplication. Comparative examination of closely related phage genomes will provide new insight unavailable from comparison of distantly related phages.

Acknowledgements

This work was supported by the 21C Frontier Microbial Genomics and Application Center Program and the Eco-Technopia 21 project. The first author was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund; KRF-2009-512-C00165).

References

Baird-Parker, A.C. 1994. Foods and microbiological risks. *Microbiology* 140, 687-695.

Bentley, C.A., J.A. Frost, and B. Rowe. 1996. Phage typing and drug resistance of *Shigella sonnei* isolated in England and Wales. *Epidemiol. Infect.* 116, 295-302.

Beutin, L., E. Strauch, and I. Fischer. 1999. Isolation of *Shigella sonnei* lysogenic for a bacteriophage encoding gene for production of Shiga toxin. *Lancet* 353, 1498.

Carver, T., M. Berriman, A. Tivey, C. Patel, U. Bohme, B.G. Barrell, J. Parkhill, and M.A. Rajandream. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24, 2672-2676.

Carver, T. and A. Bleasby. 2003. The design of Jemboss: a graphical user interface to EMBOSS. *Bioinformatics* 19, 1837-1843.

Casjens, S., D.A. Winn-Stapley, E.B. Gilcrease, R. Morona, C. Kuhlewein, J.E. Chua, P.A. Manning, W. Inwood, and A.J. Clark. 2004. The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. *J. Mol. Biol.* 339, 379-394.

Chibani-Chennoufi, S., C. Canchaya, A. Bruttin, and H. Brussow. 2004a. Comparative genomics of the T4-Like *Escherichia coli* phage JS98: implications for the evolution of T4 phages. *J. Bacteriol.* 186, 8276-8286.

Chibani-Chennoufi, S., J. Sidoti, A. Bruttin, M.L. Dillmann, E. Kutter, F. Qadri, S.A. Sarker, and H. Brussow. 2004b. Isolation of *Escherichia coli* bacteriophages from the stool of pediatric diarrhea patients in Bangladesh. *J. Bacteriol.* 186, 8287-8294.

Chibani-Chennoufi, S., J. Sidoti, A. Bruttin, E. Kutter, S. Sarker, and H. Brussow. 2004c. *In vitro* and *in vivo* bacteriolytic activities of *Escherichia coli* phages: implications for phage therapy. *Antimicrob. Agents Chemother.* 48, 2558-2569.

Comeau, A.M., C. Bertrand, A. Letarov, F. Tetart, and H.M. Krisch. 2007. Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology* 362, 384-396.

Delcher, A.L., D. Harmon, S. Kasif, O. White, and S.L. Salzberg. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27, 4636-4641.

Desplats, C. and H.M. Krisch. 2003. The diversity and evolution of the T4-type bacteriophages. *Res. Microbiol.* 154, 259-267.

Hatfull, G.F. 2001. Microbiology. The great escape. *Science* 292, 2263-2264.

Haynes, J.A. and F.A. Eiserling. 1996. Modulation of bacteriophage T4 capsid size. *Virology* 221, 67-77.

Hendrix, R.W. 2003. Bacteriophage genomics. *Curr. Opin. Microbiol.* 6, 506-511.

Herlemann, D.P., O. Geissinger, W. Ikeda-Ohtsubo, V. Kunin, H. Sun, A. Lapidus, P. Hugenholtz, and A. Brune. 2009. Genomic analysis of "*Elusimicrobium minutum*", the first cultivated representative of the phylum "*Elusimicrobia*" (formerly termite group 1). *Appl. Environ. Microbiol.* 75, 2841-2849.

Housby, J.N. and N.H. Mann. 2009. Phage therapy. *Drug Discov. Today* 14, 536-540.

James, C.E., K.N. Stanley, H.E. Allison, H.J. Flint, C.S. Stewart, R.J. Sharp, J.R. Saunders, and A.J. McCarthy. 2001. Lytic and lysogenic infection of diverse *Escherichia coli* and *Shigella* strains with a verocytotoxigenic bacteriophage. *Appl. Environ. Microbiol.* 67, 4335-4337.

Kotloff, K.L., J.P. Winickoff, B. Ivanoff, J.D. Clemens, D.L. Swerdlow, P.J. Sansonetti, G.K. Adak, and M.M. Levine. 1999. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull. World Health Organ.* 77, 651-666.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* 5, 150-163.

Loc Carrillo, C., R.J. Atterbury, A. el-Shibiny, P.L. Connerton, E. Dillon, A. Scott, and I.F. Connerton. 2005. Bacteriophage therapy to reduce *Campylobacter jejuni* colonization of broiler chickens. *Appl. Environ. Microbiol.* 71, 6554-6563.

Lowe, T.M. and S.R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955-964.

Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, and *et al*. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.

McAuliffe. 2007. The new phage biology: from genomics to applications. *In* S. McGrath and D. van Sinderen (eds.), Bacteriophage: Genetics and Molecular Biology, Caister Academic Press, Norfolk, UK.

Merril, C.R., B. Biswas, R. Carlton, N.C. Jensen, G.J. Creed, S. Zullo, and S. Adhya. 1996. Long-circulating bacteriophage as antibacterial agents. *Proc. Natl. Acad. Sci. USA* 93, 3188-3192.

Merril, C., D. Scholl, and S. Adhya. 2006. Phage therapy, pp. 725-741. *In* R.L. Calendar and S.T. Abedon (eds.), The Bacteriophages, Oxford University Press, New York, N.Y., USA.

Miller, E.S., J.F. Heidelberg, J.A. Eisen, W.C. Nelson, A.S. Durkin, A. Ciecko, T.V. Feldblyum, and *et al*. 2003a. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J. Bacteriol.* 185, 5220-5233.

Miller, E.S., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Ruger. 2003b. Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* 67, 86-156.

Muniesa, M., R. Serra-Moreno, and J. Jofre. 2004. Free Shiga toxin bacteriophages isolated from sewage showed diversity although the *stx* genes appeared conserved. *Environ. Microbiol.* 6, 716-725.

Nolan, J.M., V. Petrov, C. Bertrand, H.M. Krisch, and J.D. Karam. 2006. Genetic diversity among five T4-like bacteriophages. *Virol. J.* 3, 30-44.

Pedulla, M.L., M.E. Ford, J.M. Houtz, T. Karthikeyan, C. Wadsworth, J.A. Lewis, D. Jacobs-Sera, and *et al*. 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell* 113, 171-182.

Penatti, M.P., L.M. Hollanda, G. Nakazato, T.A. Campos, M. Lancellotti, M. Angellini, M. Brocchi, M.M. Rocha, and W. Dias da Silveira. 2007. Epidemiological characterization of resistance and PCR typing of *Shigella flexneri* and *Shigella sonnei* strains isolated from bacillary dysentery cases in Southeast Brazil. *Braz. J. Med. Biol. Res.* 40, 249-258.

Preston, M.A. and A.A. Borczyk. 1994. Genetic variability and molecular typing of *Shigella sonnei* strains isolated in Canada. *J. Clin. Microbiol.* 32, 1427-1430.

Roh, S.W., K.H. Kim, Y.D. Nam, H.W. Chang, E.J. Park, and J.W. Bae. 2009. Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. *ISME J.* 4, 1-16.

Sambrook, J. and D.W. Russell. 2001. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, New York, N.Y., USA.

Sechter, I. and D. Cohen. 1990. Developing epidemiological markers for *Shigella sonnei*: a proposed phage-typing scheme. *Public Health Rev.* 18, 261-266.

Stafford, R., M. Kirk, C. Selvey, D. Staines, H. Smith, C. Towner, and M. Salter. 2007. An outbreak of multi-resistant *Shigella sonnei* in Australia: possible link to the outbreak of shigellosis in Denmark associated with imported baby corn from Thailand. *Euro. Surveill.* 12, E070913.

Strauch, E., R. Lurz, and L. Beutin. 2001. Characterization of a Shiga toxin-encoding temperate bacteriophage of *Shigella sonnei*. *Infect. Immun.* 69, 7588-7595.

Sulakvelidze, A. 2005. Phage therapy: an attractive option for dealing with antibiotic-resistant bacterial infections. *Drug Discov. Today* 10, 807-809.

Summers, W.C. 2001. Bacteriophage therapy. *Annu. Rev. Microbiol.* 55, 437-451.

Talukder, K.A., Z. Islam, D.K. Dutta, M.A. Islam, B.K. Khajanchi, I.J. Azmi, M.S. Iqbal, and *et al*. 2006. Antibiotic resistance and genetic diversity of *Shigella sonnei* isolated from patients with diarrhoea between 1999 and 2003 in Bangladesh. *J. Med. Microbiol.* 55, 1257-1263.

Tetart, F., C. Desplats, M. Kutateladze, C. Monod, H.W. Ackermann, and H.M. Krisch. 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. *J. Bacteriol.* 183, 358-366.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876-4882.

von Seidlein, L., D.R. Kim, M. Ali, H. Lee, X. Wang, V.D. Thiem, G. Canh do, and *et al*. 2006. A multicentre study of *Shigella diarrhoea* in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLoS Med.* 3, e353.

Watson, C. 2006. Death from multi-resistant shigelloses: a case study from Fiji. *Pac. Health Dialog.* 13, 111-114.

Weinbauer, M.G. 2004. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28, 127-181.

Wheeler, D.A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, and *et al*. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872-876.

Young, P., M. Ohman, M.Q. Xu, D.A. Shub, and B.M. Sjoberg. 1994. Intron-containing T4 bacteriophage gene *sunY* encodes an anaerobic ribonucleotide reductase. *J. Biol. Chem.* 269, 20229-20232.

Zuber, S., C. Ngom-Bru, C. Barretto, A. Bruttin, H. Brussow, and E. Denou. 2007. Genome analysis of phage JS98 defines a fourth major subgroup of T4-like phages in *Escherichia coli*. *J. Bacteriol.* 189, 8206-8214.