The Journal of Microbiology (2011) Vol. 49, No. 6, pp. 927-934
DOI 10.1007/s12275-011-1577-0

Comparative Genomic Analysis of Bacteriophage EP23 Infecting *Shigella sonnei* and *Escherichia coli*

Ho-Won Chang[1] and Kyoung-Ho Kim[2*]

[1]*Busan Regional Korea Food & Drug Administration,* [2]*Department of Microbiology, Pukyong National University, Busan 608-737, Republic of Korea*

(Received November 17, 2011 / Accepted December 7, 2011)

Bacteriophage EP23 that infects *Escherichia coli* and *Shigella sonnei* was isolated and characterized. The bacteriophage morphology was similar to members of the family *Siphoviridae*. The 44,077 bp genome was fully sequenced using 454 pyrosequencing. Comparative genomic and phylogenetic analyses showed that EP23 was most closely related to phage SO-1, which infects *Sodalis glossinidius* and phage SSL-2009a, which infects engineered *E. coli*. Genomic comparison indicated that EP23 and SO-1 were very similar with each other in terms of gene order and amino acid similarity, even though their hosts were separated in the level of genus. EP23 and SSL-2009a displayed high amino acid similarity between their genes, but there was evidence of several recombination events in SSL-2009a. The results of the comparative genomic analyses further the understanding of the evolution and relationship between EP23 and its bacteriophage relatives.

***Keywords*:** Bacteriophage EP23, *Shigella sonnei* phage, *Escherichia coli* phage, *Siphoviridae,* comparative genomic analysis

The bactericidal abilities of bacteriophages have been investigated as a method to cure bacterial diseases (Merril *et al.*, 1996; Hatfull, 2001; Summers, 2001). This so-called (bacterio) phage therapy is regarded as a supplemental or an alternative method to control bacteria, especially those resistant to several antibiotics (Sulakvelidze, 2005; Merril *et al.*, 2006; Housby and Mann, 2009). To apply bacteriophages for phage therapy, it is essential to isolate suitable bacteriophages that lyse the disease-causing strains and to understand their physiological properties. Sequencing the whole genome of the isolated phage and decoding its contents helps to understand phage's physiological characteristics. Because of their high abundance and diversity, bacteriophages have been considered as a huge bioresource pool for biotechnology efforts including the design of bacterial controlling agents used in the phage therapy (Marks and Sharp, 2000; Clark and March, 2006). However, relatively few phage genomes have been sequenced and the information deposited in public databases was limited as compared to bacterial genomes. Less than 700 phage genomes are available at the NCBI genomic database (http://www.ncbi.nlm.nih.gov/genome), in contrast to more than 1600 bacterial genomes, as of October 2011, for example. Considering the much shorter size of phage genomes, it is less time-consuming and economical to sequence phage genomes than bacterial ones. However, full sequencing of a phage genome is often difficult, because some of the phage genes have lethal effects on the cloning host, mainly *Escherichia coli* (Zuber *et al.*, 2007). Next generation sequencing technology such as 454 pyrosequencing eliminates the need of *in vivo* cloning and bypasses the lethal effect of phage genes (Zuber *et al.*, 2007).

Recently, viral metagenomic studies have investigated viral diversities of various environments including the sea (Breitbart *et al.*, 2002; Kim and Bae, 2011), soil (Kim *et al.*, 2008), food (Park *et al.*, 2011) and human feces (Breitbart *et al.*, 2003; Kim *et al.*, 2011) without isolation and cultivation of viruses. The sequences obtained from those studies showed that large proportion of sequences obtained from environmental viruses are totally unrelated to known viruses; even the metagenomic sequences showing relationships with known viruses had low similarities (Breitbart *et al.*, 2002, 2003; Kim *et al.*, 2008; Kim *et al.*, 2011; Kim and Bae, 2011; Park *et al.*, 2011). Thus, most viruses in environment have not yet been revealed sufficiently. More effort to isolate new phages and sequence their genomes must be made to accumulate sufficient viral genome data for comparative genomic analysis.

Strains of the genus *Shigella* cause shigellosis, one of the most serious endemics in the world, according to the World Health Organization. *Shigella sonnei* is one of the principal causes of shigellosis, mainly in the industrialized countries (Preston and Borczyk, 1994). Drug resistant and multi-drug resistant *Shigella* species are becoming more prevalent especially in Asian and in other countries (Bentley *et al.*, 1996; Talukder *et al.*, 2006; von Seidlein *et al.*, 2006; Penatti *et al.*, 2007; Stafford *et al.*, 2007) mainly because of the improper use of antibiotics. Several strains of *Shigella*-infecting phages have been described (Beutin *et al.*, 1999; James *et al.*, 2001; Strauch *et al.*, 2001; Muniesa *et al.*, 2004), but few genomic sequences of phages infecting *Shigella* have been determined, except for a few strains including temperate bacteriophage Sf6 that infects *S. flexneri* (Casjens *et al.*, 2004) and unpublished genomes in the GenBank database.

In the course of an earlier study to isolate *Shigella*-infecting phages, we isolated phage SP18, a T4-like phage in the family *Myoviridae* infecting *S. sonnei* and described its physiological and genomic characteristics (Kim *et al.*, 2010). In this study,

* For correspondence. E-mail: kimkh@pknu.ac.kr; Tel.: +82-51-629-5611; Fax: +82-51-629-5619

a new phage designated EP23, which was determined to infect *S. sonnei* as well as *E. coli* and belonging to the family *Siphoviridae* was isolated and its basic characteristics were described. Its genome was fully sequenced using 454 pyrosequencing and analyzed comparatively with its closely related phages.

Materials and Methods

Bacteriophage isolation

A water sample was collected from the Gap river, Daejeon, Korea (36° 26′ 14.5″ N 127° 23′ 39.4″ E) using a sterilized plastic bag and transferred immediately to laboratory. The sample was filtered through a filter paper to remove particulates and through a 0.22 μm pore membrane (Millipore Corporation, USA) to remove bacteria. The filtered water sample was added to the *Escherichia coli* KCTC 2223 culture grown in advance and the mixture was incubated to enrich for phages capable of infecting the *E. coli* strain. *E. coli* KCTC 2223 was obtained from the Korean Collection for Type Cultures (KCTC) and grown in TSB (Difco, USA) at 30°C. Residual *E. coli* cells were removed by a filtration through a 0.22 μm pore membrane, and one drop of the filtrate was inoculated on a TSB agar plate on which KCTC 2223 had been spread. A single plaque that formed after incubation at 30°C was picked and transferred to a fresh plate containing KCTC 2223. Transfer was performed at least three times. The isolated phage was designated EP23.

Purification of bacteriophages

E. coli KCTC 2223 was cultured in 1 L of TSB at 30°C to early exponential phase and inoculated with phage. Bacterial cells and debris were pelleted by centrifugation at 7,000 rpm for 30 min at 4°C after observation of lysis. The supernatant was collected and filtered through 0.22 μm pore filter membranes. The filtrate was incubated for 2 hr at 4 °C after the addition of NaCl (final concentration, 0.5 M) and polyethylene glycol 8000 (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 centrifugation as previously described (Sambrook and Russell, 2001). The fraction containing the phages was purified by dialysis incubating for 2 h against a buffer containing 1 M NaCl, 50 mM Tris-HCl (pH 8.0), and 100 mM $MgCl_2$, and then overnight against a buffer containing 100 mM NaCl, 50 mM Tris-HCl (pH 8.0), and 100 mM $MgCl_2$ in a dialysis tubing. Phage particles were concentrated using Amicon® Ultra-4 centrifugal filter devices (Millipore) at 2,500 rpm for 30 min at 4°C, and then stored at -80°C.

One-step growth curve

For determination of the one-step growth curve, KCTC 2223 culture in early exponential phase was infected with phage EP23 with a multiplicity of infection of 0.1. The culture was incubated at 25°C, and the titer of phages was determined every 5 min. Aliquots were spread on the plate with or without treatment by chloroform.

Electron microscopy

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

Host range test

Twenty one strains belonging to *E. coli*, *Shigella* and *Salmonella* strains were used to determine the host range of EP23 (Table 1). A paper disk containing a drop of purified phages was placed on the center of the TSA plate (Difco) on which each bacterial strain

Table 1. Susceptibility of related bacterial strains to EP23

Culture collection number*	Species	Susceptibility
KCTC 2223	*Escherichia coli*	+
KCTC 2293	*Escherichia coli*	+
ATCC 21278	*Escherichia coli*	+
DSM 4481^T	*Escherichia blattae*	-
DSM 4560^T	*Escherichia hermannii*	-
DSM 4564^T	*Escherichia vulneris*	-
DSM 13698^T	*Escherichia fergusonii*	-
KCTC 2009	*Shigella sonnei*	-
KCTC 2518	*Shigella sonnei*	+
DSM 5570^T	*Shigella sonnei*	+
DSM 4782^T	*Shigella flexneri*	-
DSM 7532^T	*Shigella boydii*	-
DSM 13772^T	*Salmonella bongori*	-
DSM 14848^T	*Salmonella enterica* subsp. *indica*	-
DSM 14846^T	*Salmonella enterica* subsp. *enterica*	-
DSM 9386^T	*Salmonella enterica* subsp. *arizonae*	-
DSM 14847^T	*Salmonella enterica* subsp. *diarizonae*	-
DSM 9221^T	*Salmonella enterica* subsp. *houtenae*	-
DSM 9220^T	*Salmonella enterica* subsp. *salamae*	-
ATCC 13311	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	-
ATCC 13076	*Salmonella enterica* subsp. *enterica* serovar Enteritidis	-

* KCTC, the Korean Collection for Type Cultures; DSM, the German Collection of Microorganisms and Cell Cultures; ATCC, the American Type Culture Collection

had been spread. The formation of a clear zone around the paper disk after incubation for 24 hr at 30°C was considered positive for lysis by the phage EP23.

Extraction of phage genomic DNA

Genomic DNA from SP18 was isolated from the purified phages using a phenol-chloroform extraction method as described previously (Sambrook and Russell, 2001). DNA concentration was determined in triplicate using a spectrophotometer (Nanodrop Technologies, USA).

Genome sequencing by 454 pyrosequencing

The phage genome was sequenced through 454 pyrosequencing (Margulies *et al.*, 2005) by a sequencing company (Macrogen, Korea). Briefly, genomic DNA from the phage was sheared and amplified by emulsion PCR. The sequencing by synthesis was performed using the Genome Sequencer FLX (Roche, Switzerland). Newbler software was used to assembly the sequencing reads.

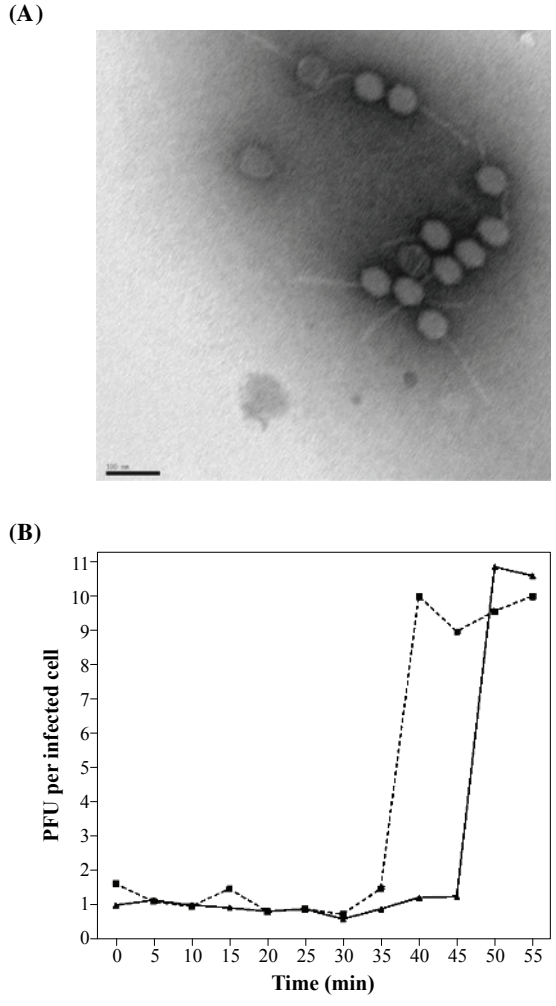

Fig. 1. Transmission electron microscopy and one-step growth curve of EP23. (A) Typical morphology of members in the family *Siphoviridae* was shown. The purified phage particles were fixed on the copper grid with a carbon-coated film and were negatively stained with 2% uranyl acetate. Bar indicates 100 nm. (B) One-step growth curve of phage EP23 on *E. coli* strain KCTC 2223 at 25°C. Shown are the plaque forming units per infected cell in chloroform-treated cultures (rectangle) and in untreated cultures (triangle) at different time points.

Annotation and comparison

Dot plot analysis was performed using the nucmer and the promer programs in the MUMmer package version 3.22 (Kurtz *et al.*, 2004). Prediction of open reading frames (ORFs) and comparison of ORFs from relative phages was performed using the RAST server (Aziz *et al.*, 2008). Three potential start codons, ATG, TTG or GTG, were used to predict ORFs. tRNA gene was predicted with the tRNAscan-SE program version 1.21 (Schattner *et al.*, 2005). BLASTP analysis against the NCBI non-redundant protein database (Jun. 2011) was used to predict the function of the ORFs (e-value $< 10^{-5}$). The Artemis and ACT programs were used for genome annotation and comparison (Carver *et al.*, 2008). Pairwise comparison of two genomes for ACT program was performed using TBLASTX.

Nucleotide sequence accession number

The genome sequence of EP23 was deposited in GenBank under the accession number JN984867.

Results and Discussion

Isolation and physiological characteristics of EP23

Phage EP23 was isolated from the Gap River, Daejeon, Korea. It produced clear, 1-3 mm-diameter plaques on tryptic soy broth (TSB) agar with *E. coli* strain KCTC 2223 after overnight incubation at 30°C. Electron microscopy of the EP23 virions revealed icosahedral heads 59±3 nm in diameter and non-contractile, filamentous tails 142±32 nm in length (Fig. 1A). The morphology was typical of the viral family *Siphoviridae* described by International Committee on Taxonomy of Viruses (ICTV, http://www.ictvonline.org). One-step growth curve showed that EP23 had an eclipse period of 35 min and the latent period of 45 min (Fig. 1B). The host range was determined using 21 strains belonging to the genera *Escherichia*, *Shigella* and *Salmonella* (Table 1). EP23 could infect all *E. coli* strains tested (n=3), but not other species in the genus *Escherichia*. Interestingly, the phage could infect two strains of *S. sonnei*, while another strain was impervious to infection. *S. sonnei,* one of the four species of the genus *Shigella*, is a cause of a diarrheal disease in industrialized countries such as Canada and the United States (Preston and Borczyk, 1994). No other strains or species belonging to *Shigella* and *Salmonella* showed any susceptibility to the EP23 phage.

Basic characteristics of EP23 genome

The length of the assembled genome sequence was 44,077 bp with a nucleotide composition of A (23%), C (27%), G (28%), and T (22%). The GC content of the phage genome was 54.4%, which was slightly higher than those of *E. coli* (50.6%) and *S. sonnei* (50.8%) acquired from the NCBI genome site (http://www.ncbi.nlm.nih.gov/genomes). A total of 57 ORFs were predicted and the genome had 92.2% of coding percentage. The average length of ORFs was 713 amino acids (in the range of 117-3,819 amino acids).

BLASTP analysis of ORFs

Fifty two among the fifty seven ORFs showed hits against

Table 2. Characteristics of predicted ORFs

ORF/ Strand	Start	Stop	size (aa)	Putative function [Nearest neighbor]	Accession #	e-value/ % aa identity	COG	PFAM name/e-value
p01/+	56	628	190	gp1 [SO-1]	YP_003344936.1	6E-107/99		-/-
p02/+	628	2013	461	terminase [SO-1]	YP_003344937.1	0.0/99		Terminase_6/3E-21
p03/+	2025	3545	506	phage structural protein [SSL-2009a]	YP_002720085.1	0.0/99		Pfam-B_1211/1E-49
p04/+	3565	4632	355	phage head morphogeneisis protein [SSL-2009a]	YP_002720086.1	0.0/97		Lant_dehyd_C/5E-02
p05/+	4625	4954	109	EpSSL_gp48 [SSL-2009a]	YP_002720087.1	8E-44/95		ThiC/5E-02
p06/+	4908	5066	52	gp6 [SO-1]	YP_003344941.1	2E-18/83		-/-
p07/-	5844	5080	254	EpSSL_gp50 [SSL-2009a]	YP_002720089.1	6E-144/97		-/-
p08/-	6438	5926	170	DNA polymerase III beta subunit [SO-1]	YP_003344942.1	6E-84/91		DNA_pol3_beta_2/3E-03
p09/+	6587	7291	234	gp8 [SO-1]	YP_003344943.1	1E-130/100		-/-
p10/+	7389	8483	364	phage structural protein [SO-1]	YP_003344944.1	0.0/96		-/-
p11/+	8542	9042	166	EpSSL_gp33 [SSL-2009a]	YP_002720072.1	1E-87/96		DUF1353/3E-02
p12/+	9044	9160	38	None	-	-/-		Trypan_PARP/2E-02
p13/+	9147	9500	117	phage structural protein [SSL-2009a]	YP_002720071.1	2E-60/97		Pfam-B_14119/7E-05
p14/+	9502	10095	197	putative tail protein [SSL-2009a]	YP_002720070.1	1E-109/99		-/-
p15/+	10095	10517	140	gp13 [SO-1]	YP_003344948.1	3E-72/93		-/-
p16/+	10579	11304	241	phage tail protein [SSL-2009a]	YP_002720068.1	1E-127/98	N	Phage_tail_3/4E-33
p17/+	11374	11736	120	gp15 [SO-1]	YP_003344950.1	6E-65/100		DUF1789/3E-21
p18/+	12447	15089	880	Phage tail length tape-measure protein 1 [SO-1]	YP_003344952.1	0.0/97	S	TMP_2/4E-39
p19/+	15126	15725	199	phage minor tail protein [SO-1]	YP_003344953.1	4E-111/98	S	Phage_min_tail/1E-16
p20/+	15722	16510	262	Phage minor tail protein L [SO-1]	YP_003344954.1	3E-152/99	S	Phage_tail_L/1E-40
p21/+	16513	17253	246	Phage tail assembly protein [SSL-2009a]	YP_002720062.1	1E-140/97	M	NLPC_P60/3E-09
p22/+	17244	17861	205	Phage tail assembly protein I [SO-1]	YP_003344956.1	1E-113/100	S	Lambda_tail_I/3E-52
p23/+	17858	21676	1272	Phage tail fiber protein [SO-1]	YP_003344957.1	0.0/95	S	DUF1983/2E-19
p24/+	22032	23660	542	putative tail fiber protein [SO-1]	YP_003344960.1	2E-97/71		-/-
p25/-	24261	23686	191	gp26 [SO-1]	YP_003344961.1	3E-101/93		-/-
p26/-	24858	24304	184	conserved hypothetical protein [SO-1]	YP_003344962.1	5E-72/84		Pfam-B_4365/7E-07
p27/-	25392	24922	156	gp28 [SO-1]	YP_003344963.1	1E-86/98		Acid_phosphat_B/3E-05
p28/-	25613	25389	74	None	-	-/-		RBD/1E-02
p29/-	26079	25594	161	DNA methyltransferase [SO-1]	YP_003344964.1	8E-89/97		Dam/3E-36
p30/-	26524	26066	152	gp30 [SO-1]	YP_003344965.1	1E-68/84		-/-
p31/-	26922	26605	105	gp31 [SO-1]	YP_003344966.1	3E-52/94		-/-
p32/-	28351	26924	475	helicase [SSL-2009a]	YP_002720041.1	0.0/98	K	SNF2_N/3E-09
p33/-	28574	28356	72	EpSSL_gp03 [SSL-2009a]	YP_002720042.1	6E-35/94		Arif-1/2E-02
p34/-	28791	28618	57	gp35 [SO-1]	YP_003344970.1	1E-24/100		Stc1/3E-02
p35/-	29068	28832	78	gp36 [SO-1]	YP_003344971.1	3E-36/97		VRR_NUC/3E-14
p36/-	32298	29110	1062	DNA polymerase family A [SO-1]	YP_003344973.1	0.0/98		DNA_pol_A/6E-07
p37/-	32465	32295	56	gp40 [SO-1]	YP_003344975.1	4E-23/98		-/-
p38/-	33198	32524	224	gp41 [SO-1]	YP_003344976.1	2E-96/96		DUF2815/6E-42
p39/-	33531	33289	80	gp42 [SO-1]	YP_003344977.1	1E-37/95		-/-
p40/-	34954	33524	476	gp43 [SO-1]	YP_003344978.1	0.0/98		DUF2800/1E-87
p41/-	35133	34954	59	None	-	-/-		-/-
p42/-	35384	35133	83	EpSSL_gp13 [SSL-2009a]	YP_002720052.1	6E-41/100		-/-
p43/-	35558	35433	41	None	-	-/-		DUF3572/3E-02
p44/-	36078	35563	171	gp46 [SO-1]	YP_003344981.1	1E-37/80		Pfam-B_3549/8E-04
p45/+	36481	36810	109	gp47 [SO-1]	YP_003344982.1	5E-54/96		Sigma70_r4_2/1E-05
p46/+	36821	39085	754	putative helicase-primase [SSL-2009a]	YP_002720055.1	0.0/97		Pfam-B_494/5E-13
p47/-	39497	39291	68	None	-	-/-		DUF1572/1E-01
p48/+	39549	39875	108	gp50 [SO-1]	YP_003344985.1	8E-27/71		MM_CoA_mutase/2E-02
p49/+	39975	40154	59	gp51 [SO-1]	YP_003344986.1	5E-26/95		-/-
p50/+	40151	40693	180	gp52 [SO-1]	YP_003344987.1	8E-74/79		-/-
p51/+	40901	41200	99	gp53 [SO-1]	YP_003344988.1	9E-42/86		-/-
p52/+	41253	41543	96	putative holin-like class II protein [SO-1]	YP_003344989.1	2E-45/96		-/-
p53/+	41540	41785	81	putative holin-like class I protein [SO-1]	YP_003344990.1	2E-36/95		-/-
p54/+	41772	42263	163	lysozyme [SO-1]	YP_003344991.1	2E-86/95	R	Phage_lysozyme/5E-15
p55/+	42280	42447	55	gp57 [SO-1]	YP_003344992.1	9E-14/93		-/-
p56/+	42685	43827	380	phosphoesterase [SSL-2009a]	YP_002720081.1	0.0/98		Metallophos_2/3E-08
p57/+	43824	44036	70	gp59 [SO-1]	YP_003344994.1	8E-27/89		DUF1515/2E-02

non-redundant protein databases of NCBI (e-value$<10^{-5}$). Thirty eight ORFs showed the highest bit-score with *Sodalis* phage SO-1 (GQ502199, unpublished) and the other fourteen ORFs with *E. coli* phage SSL-2009a (Li *et al.*, 2011). Four among five ORFs that showed no hits in the above analyses had hits against the PFAM protein family database, although the e-value was not low (Table 2). Table 2 also summarizes basic characteristics of predicted ORFs. The comparison of

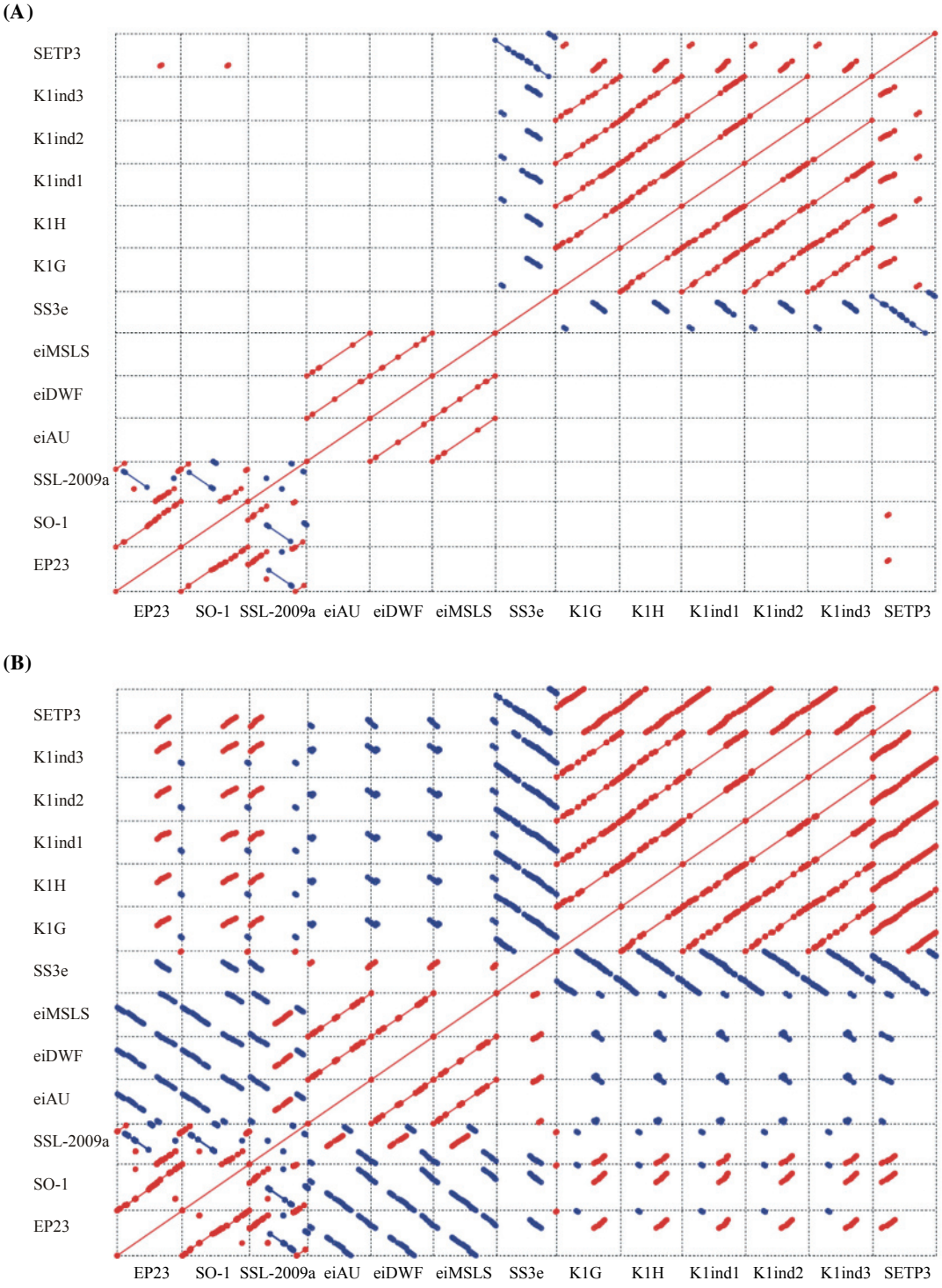

Fig. 2. Dot plot comparison between 13 related phages including EP23, based on (A) nucleotide vs nucleotide and (B) amino acid vs amino acid comparisons. Forward matches are shown in red and reverse matches are shown in blue. SSL-2009a showed the mixed orientation of genomes.

ORFs with the individual closest phages showed that 51 ORFs of EP23 had hits with SO-1 with average amino acid identity of 93% and average coverage of 97%. In the case of SSL-2009a, 40 ORFs of EP23 showed hits with average amino acid identity of 95% and coverage of 89%. Although the orders in the list of hits were behind those of phages and the identities were lower than those of phages, a large number of hits were related to bacteria. The conserved ORFs such as genes coding terminase, phage structural proteins, DNA polymerases, DNA methyltransferase, helicase, lysozyme, and phosphoesterase showing many hits with phages also showed many hits with bacteria. A large portion of those hits might be originated from prophages or phage-derived genes in bacterial genomes.

Dot plot analysis of genomes of related phages

The relationship of 13 phages showing hits with EP23 from BLASTP analysis was analyzed with dot plot comparison of their genomes based on nucleotides or amino acid sequences (Fig. 2A, B). The nucleotide-based dot plot showed clearly that 13 phages were classified into 3 groups, as shown in Fig. 2A. EP23, SO-1, and SSL-2009a were composed of a closely-related group. The amino acid dot plot showed that all phages had conserved regions in the parts of their genomes with other phages. The phages could be classified into two groups including the group composed of EP23, SO-1, SSL-2009a, eiAU, EiDWF, and eiMSLS and the group composed of the others. When forward (red color) and reverse matches (blue color) were shown at a single genome vs genome comparison, the results implied that inversion of genome fragments have occurred between two genomes after diversification from a common ancestor. SSL-2009a showed different and complex gene order with other phages (Figs. 2B and 3A). Long reverse matches shown in the comparison with EP23 and eiAU-related phages are simply because different ordering (reverse complement or not) of genome sequences. The sequence of EP23 genome was adjusted to follow the same direction of the SO-1 phage, the closest relative.

Genome structures of EP23 and related phages

Phage structural proteins determining head and tail structures were conserved in the related phages including SO-1, SSL-2009a, and eiAU phages. The order of phage structural genes of EP23 were also conserved in SO-1 and eiAU phages, but not in the case of SSL-2009a, where phosphoesterase - terminase - head morphogenesis proteins were inverted (Fig. 3B).

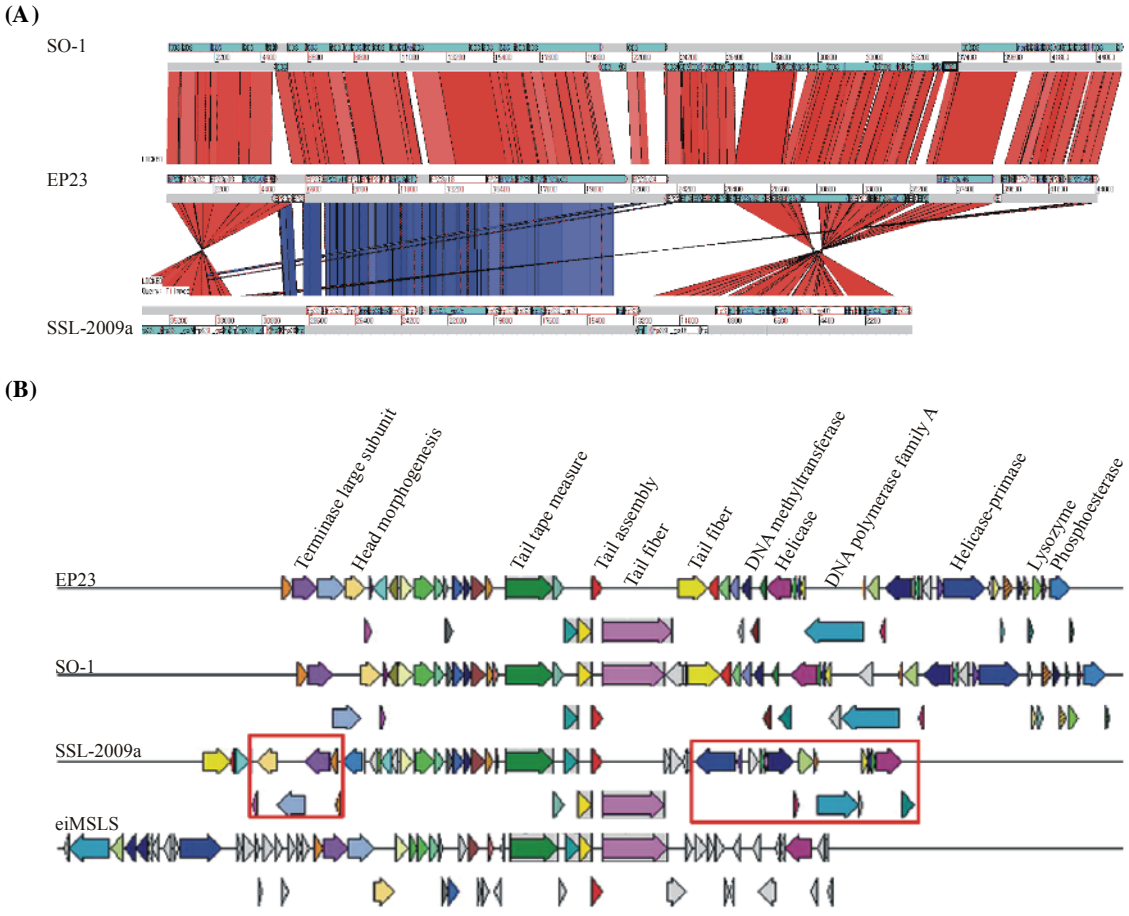

Fig. 3. Genome comparison (A) between SO-1 and SSL-2009a using the ACT program and (B) between SO-1, SSL-2009a and eiAU using the RAST server. (A) The genome of SSL-2009a is flipped to show the clearer view. (B) ORFs with similar sequences are grouped with the same color. Red boxes are the genomic regions where inversion of genes was observed.

SSL-2009a had another inverted region including helicase - DNA polymerase family A - helicase-coding genes (Fig. 3B). The conserved order of the phage morphogenetic genes was reported in temperate phages (Casjens, 2003). ORF1 and ORF2 may encode a small and a large subunit of terminase, respectively, according to a previous study of eiAU phages (Carrias *et al.*, 2011). The terminase small subunit of SSL-2009a was split into two ORFs, when considering that ORF1 had high similarity (>90%) with ORF43 and ORF44 of the phage SSL-2009a in the front and rear part of the amino acid sequence, respectively. An interesting observation was evident between EP23 and SSL-2009a. Two phages had high similarities in amino acid sequences (95.5% on average). But, their gene orders were not conserved with each other. Several inversions were observed as revealed in the dot plot analysis and the comparison of gene orders. The recombination seems to have been occurred in high frequency in the SSL-2009a rather than EP23, because the genome coordination of EP23 was also found in other related phages, as shown in Figs. 2A and 2B. Considering the high similarity in amino acid sequences, it is unusual to observe high frequency of recombination in SSL-2009a.

Host difference and phage diversity

Even if EP23 had a closely related genome sequence with SO-1, their hosts were different at the genus level. The SO-1 phage infects *Sodalis glossinidius* strain GA-SG according to the description of its genome sequence (GenBank accession no. GQ502199). The strain had 94.6% and 94.1% 16S rRNA gene sequence similarities with *E. coli* (EU014689) and *S. sonnei* (AB273732), respectively, the hosts of EP23. Although the hosts of SO-1 and EP23 were distantly related with each other, their amino acid sequence similarities were high (average similarity between two phages was 93.3%). There are known to be significant barriers for bacteriophages to switch hosts (Hatfull, 2008), whereas broad-host-range bacteriophages that infect bacteria belonging to different classes, *Gammaproteobacteria, Alphaproteobacteria*, and *Betaproteobacteria* were also reported (Jensen *et al.*, 1998). Concerning the broad host range of single phage in the level of class, it is plausible to find high similarities between two phages that have hosts different in the level of genus. Two explanations are possible. One is that they might have common hosts which are not determined yet by experiments. Another possibility is that one of phages had overcome the barrier to switch hosts recently without significant differentiation of genome sequences.

Virulent characteristics of EP23

The relatives of EP23, SSL-2009a, and eiAU phages have been reported as virulent phages (Carrias *et al.*, 2011; Li *et al.*, 2011). EP23 can be assigned into virulent phages because of its lytic ability, the presence of lysis genes such as endolysin and holin, and no similarities to lysogenic genes coding integrase, repressor, and anti-repressor proteins (Carrias *et al.*, 2011). None of the ORFs showed any similarities with pathogenecity factors that can cause problems in clinical usage. Virulent characteristics and no possible pathogen factors make it feasible to be a potential candidate for therapeutic application.

Although mosaicism of viral genomes is a well-known evolutionary mechanism of viruses, vertical evolutionary mechanisms were also observed and postulated in phages (Brussow and Desiere, 2001). Virulent phages have a tendency not to exchange DNA beyond the boundary of their group unlike temperate phages (Chopin *et al.*, 2001). Genomes of closely related phages with different origin of isolation can provide a key to understand short-term evolutionary mechanisms of phages. In this study, the phages with high similarities in amino acid sequences to each other showed different genomic aspects in terms of topology of genomic organization and hosts. This result might imply that topological rearrangement of genomes and change of hosts have lower barriers than changes of amino acid sequences during evolution. This observation sheds light on the mechanisms of vertical evolutionary changes of phage genomes.

Acknowledgements

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF 2011-0014901).

References

Aziz, R.K., D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, and *et al.* 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics.* 9, 75.

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.

Breitbart, M., I. Hewson, B. Felts, J.M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer. 2003. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* 185, 6220-6223.

Breitbart, M., P. Salamon, B. Andresen, J.M. Mahaffy, A.M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. USA* 99, 14250-14255.

Brussow, H. and F. Desiere. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. *Mol. Microbiol.* 39, 213-222.

Carrias, A., T.J. Welch, G.C. Waldbieser, D.A. Mead, J.S. Terhune, and M.R. Liles. 2011. Comparative genomic analysis of bacteriophages specific to the channel catfish pathogen *Edwardsiella ictaluri*. *Virol. J.* 8, 6.

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.

Casjens, S. 2003. Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* 49, 277-300.

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.

Chopin, A., A. Bolotin, A. Sorokin, S.D. Ehrlich, and M. Chopin. 2001. Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res.* 29, 644-651.

Clark, J.R. and J.B. March. 2006. Bacteriophages and biotechnology: vaccines, gene therapy and antibacterials. *Trends Biotechnol.* 24,

212-218.
Hatfull, G.F. 2001. Microbiology. The great escape. *Science* 292, 2263-2264.
Hatfull, G.F. 2008. Bacteriophage genomics. *Curr. Opin. Microbiol.* 11, 447-453.
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.
Jensen, E.C., H.S. Schrader, B. Rieland, T.L. Thompson, K.W. Lee, K.W. Nickerson, and T.A. Kokjohn. 1998. Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.* 64, 575-580.
Kim, K.H. and J.W. Bae. 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Environ. Microbiol.* 77, 7663-7668.
Kim, K.H., H.W. Chang, Y.D. Nam, S.W. Roh, and J.W. Bae. 2010. Phenotypic characterization and genomic analysis of the *Shigella sonnei* bacteriophage SP18. *J. Microbiol.* 48, 213-222.
Kim, K.H., H.W. Chang, Y.D. Nam, S.W. Roh, M.S. Kim, Y. Sung, C.O. Jeon, H.M. Oh, and J.W. Bae. 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl. Environ. Microbiol.* 74, 5975-5985.
Kim, M.S., E.J. Park, S.W. Roh, and J.W. Bae. 2011. Diversity and abundance of single-stranded DNA viruses in human faeces. *Appl. Environ. Microbiol.* 77, 8062-8070.
Kurtz, S., A. Phillippy, A.L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S.L. Salzberg. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12.
Li, S., L. Liu, J. Zhu, L. Zou, M. Li, Y. Cong, X. Rao, X. Hu, Y. Zhou, Z. Chen, and F. Hu. 2011. Characterization and genome sequencing of a novel coliphage isolated from engineered *Escherichia coli*. *Intervirology* 53, 211-220.
Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, and *et al.* 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.
Marks, T. and R. Sharp. 2000. Bacteriophages and biotechnology: a review. *J. Chem. Technol. Biotechnol.* 75, 6-17.
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, p. 725-741. *In* R. Calendar (ed.), The Bacteriophages, Oxford University Press, New York, N.Y., USA.
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.
Park, E.J., K.H. Kim, G.C. Abell, M.S. Kim, S.W. Roh, and J.W. Bae. 2011. Metagenomic analysis of the viral communities in fermented foods. *Appl. Environ. Microbiol.* 77, 1284-1291.
Penatti, M.P., L.M. Hollanda, G. Nakazato, T.A. Campos, M. Lancellotti, M. Angellini, M. Brocchi, M.M. Rocha, and W.D. 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.
Sambrook, J. and D.W. Russell. 2001. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, New York, N.Y., USA.
Schattner, P., A.N. Brooks, and T.M. Lowe. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W686-689.
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 070911.
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, M.A. Hossain, A.S. Faruque, G.B. Nair, and D.A. Sack. 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.
von Seidlein, L., D.R. Kim, M. Ali, H. Lee, X. Wang, V.D. Thiem, G. Canh do, W. Chaicumpa, M.D. Agtini, A. Hossain, and *et al.* 2006. A multicentre study of *Shigella diarrhoea* in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLoS Med.* 3, e353.
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.