# Complete genome analysis of a newly isolated Shigella sonnei phage vB_SsoM_Z31

**Bingdong Wei** ( weibingdong@dlut.edu.cn )
Jilin Academy of Agricultural Sciences
**Cong Cong**
Dalian University of Technology
**Yongping Xu**
Dalian University of Technology
**Lichun Zhang**
Jilin Academy of Agricultural Sciences
**Ling Zhen**
Jilin Academy of Agricultural Sciences
**Long Chen**
Jilin Academy of Agricultural Sciences
**Wei Yu**
Jilin Academy of Agricultural Sciences

**Research Article**

**Keywords:** Shigella sonnei, Shigella dysenteriae, Escherichia coli, ORFs, tRNAs, Felixounavirus

**Posted Date:** March 1st, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-272649/v1



**Version of Record:** A version of this preprint was published at Archives of Virology on June 11th, 2021.
See the published version at https://doi.org/10.1007/s00705-021-05121-y.

# Abstract

This work describes the characterization and genome annotation of a newly isolated lytic phage vB_SsoM_Z31 (referred to as Z31), isolated from wastewater samples collected in Dalian, China. Transmission electron microscope revealed that phage Z31 belongs to the family Myoviridae, order Caudovirales. This phage specifically infects the *Shigella sonnei*, *Shigella dysenteriae* and *Escherichia coli.* The genome of the phage Z31 is an 89,355 bp length dsDNA molecule with a G + C content of 38.87%. It has been predicted to contain 133 ORFs, and 24 tRNAs. No homologs of virulence factors or antimicrobial resistance genes were found in this phage. Based on the results of nucleotide sequence alignment and phylogenetic analysis, phage Z31 was assigned to the genus *Felixounavirus*, subfamily Ounavirnae.


# Introduction

*Shigella* species are Gram-negative, nonmotile bacilli belonging to the family Enterobacteriaceae. The genus *Shigella* includes four species: *Shigella dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei*. Shigellosis continues to be a major cause of morbidity and mortality in developing countries and is the most important cause of bloody diarrhea worldwide [1, 2]. The World Health Organization (WHO) reported that *Shigella* spp. are responsible for an estimated 165 million cases of bacillary dysentery, among which more than 100 million occur in developing countries, causing 1 million deaths annually [3]. The highest rate of *Shigella* infection (69% of cases) and the highest death rate (61% of deaths) occur in individuals younger than 5 years [4]. Among the four species of *Shigella*, *S. sonnei* is the most common cause of shigellosis in industrialized regions in Europe, North America, and Australia. Its occurrence is currently expanding in middle-income countries across Asia, Latin America, and the Middle East [5]. *Shigella* is transmitted by direct contact with an infected person, or eating contaminated food or drinking contaminated water. A wide variety of foods frequently become contaminated with *Shigella*, including fresh fruits [6], vegetables [7], ready-to-eat foods [8] and meat products [9]. Antibiotics have been used to reduce shigellosis duration; however, the gradual emergence of multidrug-resistant *Shigella* spp. has been reported in the last decade [10–12]. Thus, there is an urgent need to develop a new strategy to control, inhibit and eliminate *Shigella* spp. [13]. Bacteriophages are natural predators of bacteria, that generally kill a single bacterial stain or subtype of bacteria with high specificity. Bacteriophages have been demonstrated to potentially act as ideal antibacterial drugs. In this study, we have sequenced and analyzed the complete genome of a newly isolated *S. sonnei* phage.

# Materials And Methods

### Bacterial strains and growth condition

The bacterial strains used in this study are listed in **Table 1**. Host bacteria (*S. sonnei* CGMCC 21535) was purchased from the China General Microbiological Culture Collection Center (CGMCC). All strains were cultured in liquid LB or plated on solid LB medium with 1.5% agar. The liquid cultures were grown with

aeration at 37°C in a shaking incubator (180 rpm). The plates with solid medium were incubated at 37°C for 8-12h. The phage infection processes were studied at 37°C, under aerobic conditions in a shaking incubator (180 rpm). All strains were stocked in LB containing 50% glycerol and stored at -80°C.

## Phage isolation

Phage vB_SsoM_Z31 (referred to as Z31) was isolated from sewages according to procedures described previously by Zhang et al [14]. Water samples were collected from the 2nd Hospital of Dalian Medical University in China.

## Host range investigation and efficiency of plating analysis

Twenty three bacterial strains, including *Shigella*, *Escherichia coli, and Salmonella***(Table1)** were used to determine the lytic capacity of phage Z31 using the spot test method on the basis of its ability to form lysis zone on lawn cultures of different strains [15]. The purified phage Z31 suspension (10 μl, $10^9$ PFU/ml) was spotted directly onto the surface of a bacterial lawn culture plate and incubated overnight at 37°C. The plates were examined for the appearance of clear zones around the phage drop. Efficiency of plating (EOP) was used to evaluated the host spectra of the phage against a variety of bacterial strains (positive spot test). EOP was calculated based on phage titer on the test strains versus the phage titer on the host bacteria. The value obtained with the host strain was considered as EOP=1.

## Transmission electron microscope

The purified phage suspension ($10^9$ PFU/ml) was absorbed onto carbon coated copper grids for 10 min. The grids were then negatively stained with 2% (w/v) uranyl acetate, followed by examination using a JEM-2100EX transmission electron microscope (TEM) (JEOL CO., Tokyo, Japan) [16].

## Phage DNA purification and sequencing

Phage genomic DNA was extracted from a preparation with a high titer of phage particles ($10^{10}$ PFU/ml) using the phenol-chloroform-isoamyl alcohol method as described by Sambrook et al [17]. A DNA library was constructed according to the protocol of the Illumina TruSeq™ Nano DNA Sample Prep Kit. Whole-genome sequencing was performed by Shanghai Biozeron Biotechnology Co., Ltd. (Shanghai, China.) using the Illumina NovaSeq 6000 sequencing platform (150 bp×2) with paired-end reads. A total of 266 Mb of sequence data were obtained. The average read length was 343 bp. Low quality (Q-value<20, 97.93%) reads were filtered out using Trimmomatic v 0.36, with an approximately 2596× depth of coverage among the 773,432 reads. ABySS (http://www.bcgsc.ca/platform/bioinfo/software/abyss) was used to perform genome assembly with multiple-kmer parameters to obtain optimal results for the assembly. GapCloser software (https://sourceforge.net/projects/soapdenovo2/files/GapCloser/) was subsequently applied to fill the remaining local inner gaps and correct single-base polymorphism for the final assembly results.

### Genome analysis

Open reading frames (ORFs) were identified using the GeneMark Server (http://topaz.gatech.edu/GeneMark/genemarks.cgi) and the RAST server (http://rast.nmpdr.org/rast.cgi). The final assembled genome sequence was used to search the current protein and nucleotide databases (http://www.ncbi.nlm.nih.gov/) using the Basic Local Alignment Search Tool (BLAST). BLASTp (https://blast.ncbi.nlm.nih.gov/Blast) was used to identify the putative functions of the encoded proteins. BLASTn (https://blast.ncbi.nlm.nih.gov/Blast) was used to compare phage genome sequence similarity. Putative tRNA-encoding genes were predicted using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) [18]. The ResFinder server [19] (https://cge.cbs.dtu.dk/services/ResFinder/) and Virulence Factor Predictor [20] (https://cge.cbs.dtu.dk/services/VirulenceFinder/) were used to identify antimicrobial resistance determinants and potential virulence factors, respectively, in the Z31 genome.

The DNA polymerase and large subunit terminase sequences were employed to determine the phylogenetic position and DNA packaging strategies of the phage. The DNA polymerase and large subunit terminase amino acid sequences obtained in this study and those of other phages were selected for multiple alignments using the Clustal W algorithm, and phylogenetic trees were constructed using MEGA7 with the neighbor-joining method. The comparative analyses of phage complete genome sequences, between Z31 and the other members of the genus *Felixounavirus*, were conducted using Easyfig [21].

## Results And Discussion

One lytic phage against *S. sonnei* (CGMCC 21535) was isolated from sewage. We named this phage vB_SsoM_Z31 (referred to as Z31). TEM analysis revealed that Z31 has an icosahedral head (60±2 nm) connected to a tail (150±2 nm). Based on these structural features, Z31 was designated as a member of the Myoviridae family, Caudovirales order (**Fig.S1**). Spot test indicated that five *Shigella* strains tested and Enterotoxigenic *E. coli* K88 CVCC 83902 were lysed by the phage Z31 **(Table 1)**. In addition, EOP results revealed that phage Z31 was showed high infection efficiency against four *Shigella* strains (*S. sonnei* BNCC 192105 EOP=0.98, *S.sonnei* BNCC 108852 EOP=0.88, *S. dysenteriae* CGMCC 10983 EOP=0.68, and *S. dysenteriae* BNCC 103609 EOP=0.94) and medium infection efficiency against Enterotoxigenic *E. coli* K88 (EOP=0.41).

The results of the present study showed that phage Z31 has a double-stranded DNA genome with a length of 89,355 bp and an overall G+C content of 38.87%. Using the RAST server, we identified 133 ORFs and predicted 100 putative protein-coding genes in the genome, 33 of which were functionally assigned. Based on bioinformatic predictions, these ORFs were categorized into four functional modules, including phage structure, host lysis, phage DNA packaging and replication and hypothetical protein (**Table S2**). Using tRNAscan-SE, Z31 was found to contain 24 predicted tRNAs (**Table S3**), located between positions 73,864-79,210 bp. tRNA genes are universally distributed in dsDNA phages, and virulent phages contain more tRNAs than temperate phages, with higher codon usage bias [22]. It may be possible that phage-

encoded tRNAs enhance translation or compensate for less abundant tRNAs in the host. The large number of tRNAs might enable phages to be translated more efficiently, reduce their latency time and increase their reproduction rate [23].

Most of the ORFs of Z31 start with an AUG codon (128 ORFs, 96.2%), three start with GUG (2.3%), and two start with UUG (1.5%). The three stop codons were present in different proportions, with UAA being the most common (88 ORFs, 66.2%), followed by UGA (37 ORFs, 27.8%) and UAG (8 ORFs, 6%).

The phylogenetic tree constructed using the DNA polymerase sequence (ORF65) revealed that Z31 was most closely related to three phages of the genus *Felixounavirus*, namely vB_SpuM_SP116 (YP_009146339.1), HY02 (YP_009204997.1), and Felix O1 (AAQ14704.1), and these phages belonged to a cluster that was clearly distinct from those containing members of other genera of the subfamily Ounavirinae, family Myoviridae (Fig.2a). The DNA packaging strategies of tailed dsDNA phages can be classified into 6 types (17 subtypes): (a) cohesive ends (5' cos, lambda P2; 3' cos, HK97); (b) headful packaging (P2, P22, Sf6, T4, 933 W, phiPLPE, phiKZ); (c) host ends (Mu, D3112); (d) short direct terminal repeats (DTRs) (T7, N4, C-st); (e) long DTRs (SPO1); and (f) covalently bound terminal proteins (*Bacillus subtilis* phage ϕ29) [24, 25]. As shown in Fig. 2b, the large terminase subunits of 16 phages (genus *Felixounavirus*) formed a branch; however, these phages were separated from the other known phage groups, indicating that genus *Felixounavirus* phages may use a novel genome packaging strategy that differs from these known strategies.

The genome sequence of phage Z31 was compared to the sequences of these phages using BLASTn and BLASTp. Phage Z31 was similar to *Enterobacteria* phage KhF1(query cover, 93%; identity, 96%), *Enterobacteria* phage KhF3 (query cover, 92%; identity, 95.91%), Enterobacteria phage XTG1 (query cover, 94%; identity, 95.75%), and *Escherichia* phage vB_EcoM_LMP25 (query cover, 91%; identity, 95.97%), and *Escherichia* phage vB_EcoM_AYO145A (query cover, 90%; identity, 96.52%), and all six phages were members of the genus *Felixounavirus*. The results of genome comparison shared the most similarities with vB_EcoM_AYO145A, JK55, XTG1, KhF3, KhF1 and vB_EcoM_LMP25, starting at the same gene position and orientation (Fig.3).

BLASTp analysis revealed that most of the putative proteins of Z31 show a high degree of similarity to putative proteins of *Salmonella* phage vB_SPuM_SP116 (18/133, 13.5%), *Escherichia* phage wV8 (11/133, 8.3%), Enterobacteria phage UAB_Phi87 (10/133, 7.5%), Enterobacteria phage vB_EcoM_IME338 (9/133, 6.8%), *Salmonella* phage BPS17W1 (8/133, 6%), and *Escherichia* phage vB_EcoM_AYO145A (7/133, 5.3%). Moreover, the structural proteins of Z31 are identical to *Salmonella* phage BPS15Q2 (minor fiber protein), *Escherichia* phage vB_EcoM-VpaE1 (tail fiber), *Escherichia coli* phage (tail fiber), Enterobacteria phage UAB_Phi87 (conserved structural protein), *Salmonella* phage FSL SP-010 (structural protein), and *Salmonella* phage vB_SPuM_SP116 (tail protein).

The horizontal transfer of phage-mediated antimicrobial resistance genes plays an important role in the evaluation of bacterial antimicrobial resistance [26]. No homologs of virulence factors (Shiga toxin genes) or antimicrobial resistance genes were found in the Z31 genome.

# Conclusion

We therefore conclude that Z31 is a newly isolated phage that can potentially be used as a therapeutic agent.

# Declarations


### Funding

This work was financially supported by the Agricultural Science and Technology Innovation Program of Jilin Province (CXGC2017JQ002).


### Nucleotide sequence accession number

The GenBank accession number for phage vB_SsoM_Z31 is MN655999.

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals by any of the authors.

# References


[1] Jakhetia R, Talukder K A, Verma N K (2013) Isolation, characterization and comparative genomics of bacteriophage SfIV: a novel serotype converting phage from *Shigella flexneri*. BMC Genomics 14:677.

[2] WHO (2005) Guidelines for the control of shigellosis, including epidemics due to *Shigella* dysenteriae type 1. WHO Document Production Services, Geneva, Switzerland.

[3] WHO (2014) Antimicrobial resistance: global report on surveillance 2014 Switzerland, Geneva: WHO Press.

[4] Ameya G, Tsalla T, Getu F, et al (2018) Antimicrobial susceptibility pattern, and associated factors of *Salmonella* and *Shigella* infections among under five children in Arba Minch, South Ethiopia. Annals of Clinical Microbiology and Antimicrobials 17:1-7.

[5] Torraca V, Holt K, Mostowy S (2020) *Shigella sonnei*. Trends in Microbiology In press: 1-2

[6] Warren B R, Parish M E, Schneider K R (2006) *Shigella* as a foodborne pathogen and current methods for detection in food. Critical Reviews in Food Technology 46:551-567.

[7] Reller M E, Nelson J M, Molbak K, et al (2006) A large, multiple-restaurant outbreak of infection with *Shigella flexneri* serotype 2a traced to tomatoes. Clinic Infectious Diseases 42:163-169.

[8] Tehrani S, Harzandi N, Jabalameli L (2018) Molecular detection of *Shigella* spp. contamination in ready-to-eat salad samples in west of Tehran. International Journal of Enteric Pathogens 6:41-44.

[9] Garedew L, Hagos Z, Zegeye B, et al (2016) The detection and antimicrobial susceptibility profile of *Shigella* isolates from meat and swab samples at butchers' shops in Gondar town, Northwest Ethiopia. Journal of Infection and Public Health 9:348-355.

[10] Duy P T, Nguyen T N T, Thuy D V, et al (2020) Commensal *Escherichia coli* are a reservoir for the transfer of XDR plasmids into epidemic fluoroquinolone-resistant *Shigella sonnei*. Nature Microbiology 5:1-9.

[11] El-Kazzaz S S, Mashaly G E, Zeid M S (2020) Multidrug resistant *Shigella* sssociated with class 1 integrase and other virulence genes as a cause of diarrhea in pediatric patients. Open Journal of Medical Microbiology 10:1-16.

[12] Strysko J, Fialkowski V, Marsh Z, et al (2019) Outbreak of multidrug-resistant *Shigella sonnei* infections in a retirement community-Vermont, October-November 2018. Morbid Mortal W 68:405-406.

[13] Shahin K, Bouzari M, Wang R, et al (2019) Prevalence and molecular characterization of multidrug-resistant *Shigella* species of food origins and their inactivation by specific lytic bacteriophages. International Journal of Food Microbiology 305:108252.

[14] Zhang W H, Mi Z Q, Yin X Y, et al (2013) Characterization of *Enterococcus faecalis* phage IME-EF1 and its endolysin. Plos One 8:e80436.

[15] Jun J W, Yun S K, Kim H J, et al (2014) Characterization and complete genome sequence of a novel N4-like bacteriophage, pSb-1 infecting *Shigella boydii*. Research in Microbiology 165:671-678.

[16] Yuksel S, Thompson K, Ellis A, et al (2001) Purification of *Piscirickettsia salmonis* and associated phage particles. Diseases of Aquatic Organisms 44:231-235.

[17] Sambrook J, Russell D W, Nina I, et al. (2001) Chapter 2: Bacteriophage λ and its vectors. In: Sambrook J, Russell DW (eds) Molecular cloning: a laboratory manual, vol 1, 3rd edn. Cold Spring Harbor, New York, pp 2.56.

[18] Schattner P, Brooks A N, Lowe T M (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Research 33:W686-W689.

[19] Zankari E, Hasman H, Cosentino S, et al (2012) Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy 67:2640-2644.

[20] Joensen K G, Scheutz F, Lund O, et al (2014) Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. Journal of Clinical Microbiology 52:1501-1510.

[21] Sullivan M. J, Petty N. K, Beatson S. A (2011) Easyfig: a genome comparison visualizer. Bioinformatics; 27 (7):1009-1010

[22] Bailly-Bechet M, Vergassola M, Rocha E (2007) Causes for the intriguing presence of tRNAs in phages. Genome Research 17:1486-1495.

[23] Santos S B, Kropinski A M, Ceyssens P J, et al (2011) Genomic and proteomic characterization of the broad-host-range *Salmonella* phage PVP-SE1: creation of a new phage genus. Journal of Virology 85:11265-11273.

[24] Casjens S R, Gilcrease E B (2009) Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. Methods Molecular Biology 502:91-111.

[25] Merrill B D, Ward A T, Grose J H, et al (2016) Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. BMC Genomics 17:679-694.

[26] Penadés J R, Chen J, Quiles-Puchalt N, et al (2015) Bacteriophage-mediated spread of bacterial virulence genes. Current Opinion in Microbiology 23:171-178.

# Tables

**Table 1 Host range of the phage Z31**

| Bacterial Species | Source and Strain [a] | Spot test [b] | EOP [c] |
|---|---|---|---|
| *Shigella sonnei* | CGMCC 21535 |  | 1 (host) |
| *Shigella sonnei* | BNCC 192105 |  | 0.98 |
| *Shigella sonnei* | BNCC 108852 |  | 0.88 |
| *Shigella dysenteriae* | CGMCC 10983 |  | 0.68 |
| *Shigella dysenteriae* | BNCC 103609 |  | 0.94 |
| *Shigella dysenteriae* | BNCC 339874 |  | NT |
| *Shigella boydii* | BNCC 186201 | /  | <0.001 |
| *Shigella flexneri* | CGMCC 10865 | /  | <0.001 |
| *Shigella flexneri* | BNCC 186377 | /  | <0.001 |
| *Shigella flexneri* | BNCC 138608 |  | NT |
| *Shigella flexneri* | BNCC 185915 |  | NT |
| *Shigella flexneri* | BNCC 337103 |  | NT |
| *Shigella flexneri* | BNCC 186377 |  | NT |
| *Shigella flexneri* | BNCC 232380 | /  | <0.001 |
| Shiga toxin-producing *Escherichia coli* | CGMCC 10668 | /  | <0.001 |
| Enterotoxigenic *Escherichia coli* K88 | CVCC 83902 |  | 0.41 |
| *Escherichia coli* | CVCC 233 |  | NT |
| *Escherichia coli* | CVCC 236 |  | NT |
| *Escherichia coli* | CVCC 238 |  | NT |
| *Escherichia coli* | BNCC 125988 |  | NT |
| *Salmonella pullorum* | CVCC 1795 |  | NT |
| *Salmonella enteritidis* | CVCC 3378 |  | NT |
| *Salmonella typhimurium* | CGMCC 50115 |  | NT |

[a] CGMCC= China General Microbiological Culture Collection Center; BNCC= Bena Culture Collection; CVCC= China Veterinary Culture Collection Center;

[b] Spot test is used to determine the lytic capacity of the phage, : clear plaque;  : no plaque;  / : unclear results;

[c] EOP is conducted only for the spot test positive strains; NT indicates EOP was not tested; EOP= phage titer on the test strain/phage titer on the host strain; EOP≥0.5 means high production efficiency, 0.1≤EOP 0.5 means medium production efficiency, 0.001≤EOP0.1 means low production efficiency, and EOP 0.001 means inefficiency of phage production.

**Table S2** Predicted functions of phage Z31

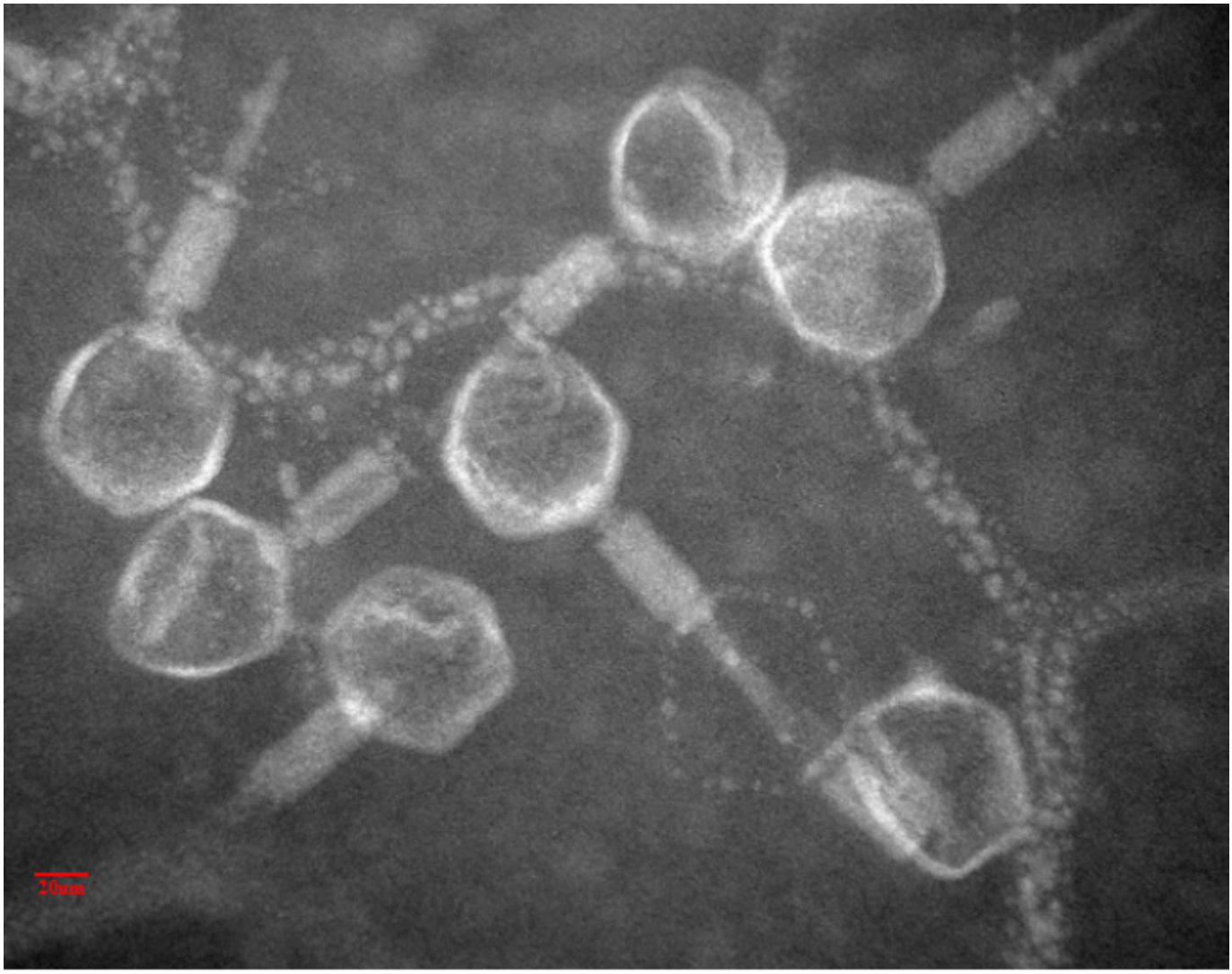
| ORF | Nucleotide position | Putative function and best match | Identity (%) | E-value | Accession number |
|---|---|---|---|---|---|
| **Phage structure** | | | | | |
| 63 | 37656-38468 | minor tail protein [*Salmonella* phage BPS15Q2] | 99 | 1.20E-134 | ANT42427.1 |
| 83 | 50525-52912 | tail fiber protein [*Escherichia* phage vB_EcoM-VpaE1] | 83 | 2.90E-282 | AIW02335.1 |
| 84 | 52959-54146 | phage tail fiber protein [Enterobacteria phage KhF2] | 100 | 0 | ANZ51996.1 |
| 97 | 63341-64693 | phage conserved structural protein [Enterobacteria phage UAB_Phi87] | 97 | 2.10E-237 | AFQ96174.1 |
| 103 | 67756-68133 | structural protein [*Salmonella* phage FSL SP-010] | 100 | 8.10E-62 | AGF88809.1 |
| 120 | 83052-83219 | putative tail protein [*Salmonella* phage vB_SPuM_SP116] | 89 | 3.60E-14 | AJT60620.1 |
| **DNA packaging** | | | | | |
| 75 | 46951-47250 | putative transcriptional regulator [*Escherichia* phage EC6] | 100 | 4.80E-49 | YP_009146330.1 |
| 104 | 68145-69491 | putative head maturation protease [*Escherichia* phage vB_EcoM_AYO145A] | 100 | 2.10E-237 | AKC04885.1 |
| 108 | 71818-73419 | terminase, large subunit [*Escherichia* phage wV8] | 100 | 0 | YP_009146297.1 |
| **DNA metabolism** | | | | | |
| 16 | 7710-7910 | putative phosphatase [*Escherichia* phage EC6] | 99 | 1.70E-146 | AFU62458.1 |
| 29 | 13147-14253 | rIIB protein [*Enterobacteria* phage vB_EcoM_IME338] | 94 | 5E-176 | AWD91288 |
| 30 | 14333-16693 | rIIA protein [*Salmonella* phage BPS17L1] | 99 | 0 | QHR67261 |
| 31 | 16728-16904 | putative membrane protein [*Salmonella* phage vB_SPuM_SP116] | 100 | 1.80E-27 | YP_009200870.1 |
| 34 | 17625-19406 | nicotinamide phosphoribosyl transferase [*Salmonella* phage BPS15Q2] | 100 | 0 | ANT42456.1 |
| 35 | 19421-19912 | HNH endonuclease [*Escherichia* phage vB_EcoM-VpaE1] | 99 | 2.50E-87 | AFU63402.1 |
| 37 | 20813-21091 | ribose-phosphate pyrophosphokinase [*Escherichia* phage vB_EcoM-VpaE1] | 99 | 3.70E-161 | AIW02388.1 |

| ORF | Nucleotide position | Putative function and best match | Identity (%) | E-value | Accession number |
|---|---|---|---|---|---|
| 42 | 22538-23014 | anaerobic NTP reductase small subunit [*Salmonella* phage BPS17W1] | 100 | 4.70E-91 | AUM59293.1 |
| 45 | 23805-25949 | anaerobic nucleoside diphosphate reductase [*Escherichia* phage vB_EcoM_AYO145A] | 100 | 0 | AKC04946.1 |
| 46 | 25998-26204 | membrane protein [*Salmonella* phage BPS17W1] | 100 | 1.20E-27 | AUM59287.1 |
| 47 | 26197-26439 | glutaredoxin [*Escherichia* phage EC6] | 100 | 1.40E-38 | AFU62428.1 |
| 48 | 26439-27512 | ribonucleoside triphosphate reductase, beta chain [*Salmonella* phage FelixO1] | 100 | 2.50E-201 | YP_002922904.1 |
| 50 | 27822-30056 | ribonucleoside triphosphate reductase alpha chain [*Salmonella* phage vB_SPuM_SP116] | 100 | 0 | YP_009168686.1 |
| 56 | 32229-33269 | putative exodeoxyribonuclease [Enterobacteria phage UAB_Phi87] | 100 | 3.20E-201 | AFQ96130.1 |
| 60 | 34664-36649 | DNA primase/helicase [*Salmonella* phage BPS15Q2] | 100 | 0 | ANT42430.1 |
| 62 | 36851-37594 | deoxynucleotide monophosphate kinase [*Escherichia* phage vB_EcoM_AYO145A] | 93 | 9.50E-123 | AKC04929.1 |
| 65 | 39120-41849 | putative DNA polymerase [*Salmonella* phage vB_SPuM_SP116] | 100 | 0 | AJT60670.1 |
| 66 | 41853-42386 | HNH endonuclease [*Salmonella* phage BPS17W1] | 100 | 1.50E-98 | AUM59266.1 |
| 70 | 43289-44392 | ATP-dependent DNA ligase [*Salmonella* phage FSL SP-010] | 100 | 5.80E-217 | AGF88846.1 |
| 79 | 48398-48943 | dihydrofolate reductase [*Salmonella* phage vB_SPuM_SP116] | 98 | 4.10E-91 | AJT60657.1 |
| 80 | 48945-49844 | thymidylate synthase [*Salmonella* phage BPS17W1] | 98 | 4.10E-91 | AUM59250.1 |
| 96 | 62879-63325 | DUF3277 family protein [*Escherichia coli*] | 100 | 9.50E-78 | YP_009146309.1 |
| 119 | 82647-83051 | immunoglobulin I-set domain protein [*Escherichia* coli O157 typing phage 1] | 99 | 3.00E-70 | AKE47248.1 |
| **Host lysis** | | | | | |
| 121 | 83219-83683 | lysin [*Escherichia* phage HY02] | 100 | 1.60E-80 | YP_009219558.1 |

**Table S3** Predicted tRNAs in phage Z31.

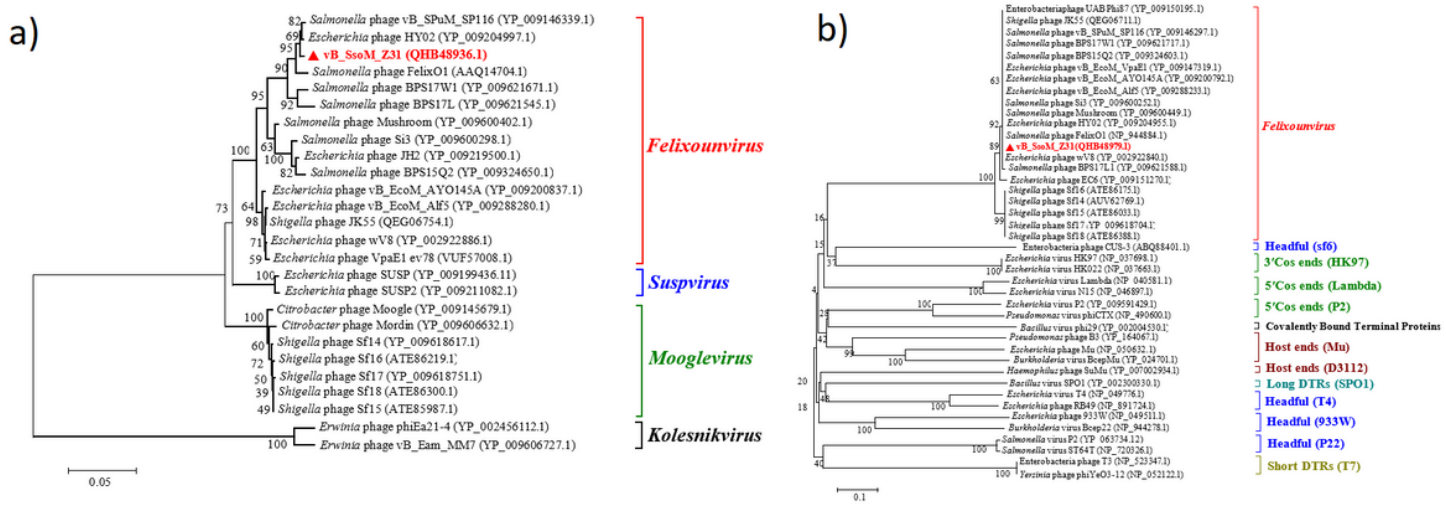
| tRNAs NO. | Begin | End | Type | Anticode | Length (bp) | Score |
|---|---|---|---|---|---|---|
| 1 | 79210 | 79137 | Pro | TGG | 74 | 57.8 |
| 2 | 79126 | 79052 | Glu | TTC | 75 | 50.1 |
| 3 | 78803 | 78730 | Asn | GTT | 74 | 40.4 |
| 4 | 78656 | 78572 | Tyr | GTA | 85 | 47.9 |
| 5 | 78562 | 78489 | Asp | GTC | 74 | 60.9 |
| 6 | 78026 | 77954 | Lys | TTT | 73 | 55.1 |
| 7 | 77946 | 77870 | Met | CAT | 77 | 47.2 |
| 8 | 77868 | 77796 | Ile | GAT | 73 | 49.9 |
| 9 | 77545 | 77469 | Arg | TCT | 77 | 59.4 |
| 10 | 77230 | 77143 | Ser | TGA | 88 | 29.8 |
| 11 | 76885 | 76811 | Leu | TAG | 75 | 39.7 |
| 12 | 76800 | 76728 | Lys | CTT | 73 | 57.2 |
| 13 | 76718 | 76646 | ALa | TGC | 73 | 47.7 |
| 14 | 76636 | 76562 | Gly | TCC | 75 | 49.4 |
| 15 | 76554 | 76481 | Thr | TGT | 74 | 49.2 |
| 16 | 76382 | 76311 | Val | TAC | 72 | 41.5 |
| 17 | 76306 | 76232 | Leu | CAA | 75 | 41.1 |
| 18 | 76114 | 76042 | Arg | ACG | 73 | 57.6 |
| 19 | 75214 | 75140 | Gln | TTG | 75 | 40.5 |
| 20 | 75137 | 75062 | Leu | TAA | 76 | 44.3 |
| 21 | 75053 | 74981 | Gln | CTG | 73 | 46.9 |
| 22 | 74946 | 74871 | His | GTG | 76 | 44.4 |
| 23 | 74856 | 74789 | Phe | GAA | 68 | 37.1 |
| 24 | 73939 | 73864 | Cys | GCA | 76 | 48.6 |

# Figures



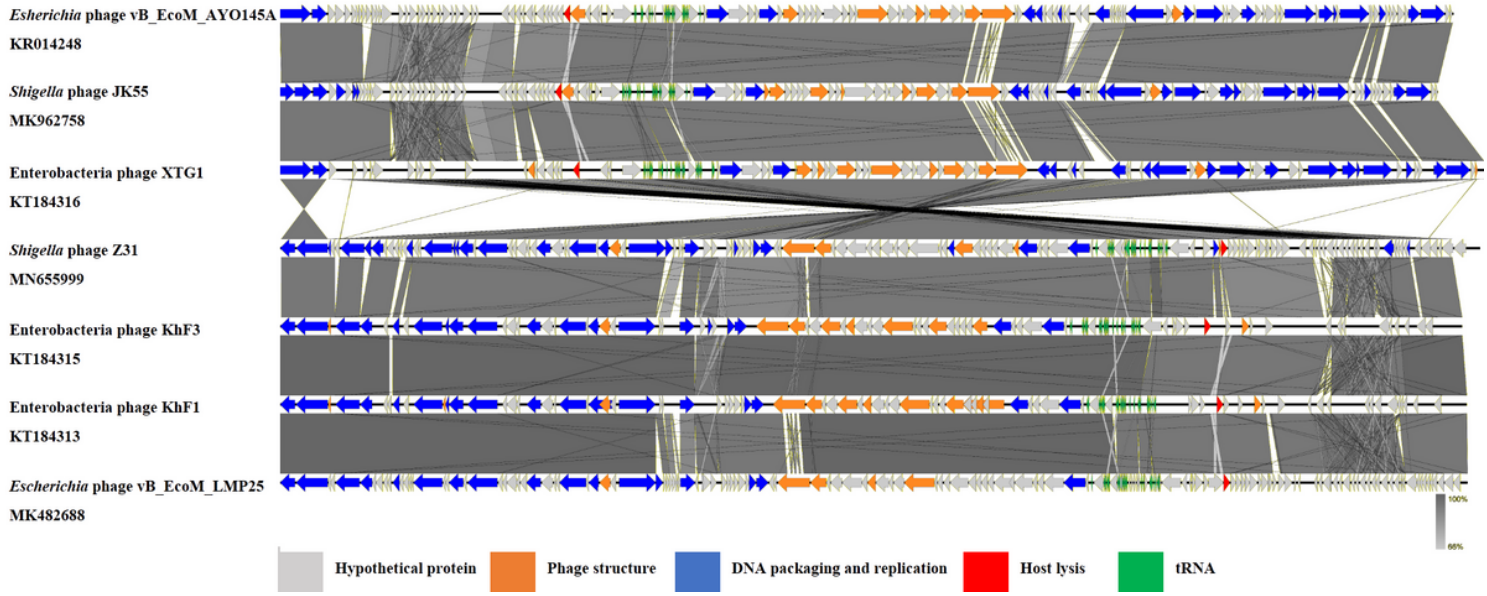

## Figure 1

Transmission electron micrograph of Z31 negatively stained with 2% (w/v) uranyl acetate. The scale bar represents 20nm.




## Figure 2

Neighbor-joining phylogenetic tree based on amino acid sequence of (a) DNA polymerase and (b) terminase large subunit, showing the relationship between phage Z31 and other phages. Values at the nodes indicate the bootstrap support calculated from 1000 replicates.




## Figure 3

Comparison of the genome of phage Z31 to other Felixounavirus members (Escherichia phage vB_EcoM_AYO145A, Shigella phage JK55, Enterobacteria phage XTG1, Enterobacteria phage KhF3, Enterobacteria phage KhF1 and Escherichia phage vB_EcoM_LMP25) using Easyfig software. The different color arrows represent CDS and tRNAs in the whole genome sequence. The direction of arrows indicated the transcription direction. The grey bars indicated the similarity of two pairs of sequences, and the intensity of grey indicated the degree of sequence similarity.