Complete genome analysis of the newly isolated Shigella sonnei phage vB_SsoM_Z31

Bingdong Wei1,2,3 · Cong Cong4 · Lichun Zhang1 · Lin Zheng1,3 · Long Chen1,3 · Wei Yu1,3 · Yongping Xu2,4

Received: 21 May 2020 / Accepted: 14 April 2021 / Published online: 11 June 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract
This work describes the characterization and genome annotation of the newly isolated lytic phage vB_SsoM_Z31 (referred to as Z31), isolated from wastewater samples collected in Dalian, China. Transmission electron microscopy revealed that phage Z31 belongs to the family Myoviridae, order Caudovirales. This phage specifically infects Shigella sonnei, Shigella dysenteriae, and Escherichia coli. The genome of the phage Z31 is an 89,355-bp-long dsDNA molecule with a G+C content of 38.87%. It was predicted to contain 133 ORFs and encode 24 tRNAs. No homologs of virulence factor genes 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 Ounavirinae.

Introduction
Shigella is a genus of Gram-negative, nonmotile bacilli belonging to the family Enterobacteriaceae. It 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) has reported that Shigella spp. are responsible for an estimated 165 million cases of bacillary dysentery, more than 100 million of which 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 of age [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, 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 shorten the duration of shigellosis, but 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, and they generally kill a single bacterial strain or subtype of bacteria with high specificity. Bacteriophages have demonstrated potential as 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 conditions
The bacterial strains used in this study are listed in Table 1. Host bacteria (S. sonnei CGMCC 21535) were 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 8-12 h. 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 sewage according to procedures described previously by Zhang et al. [14]. Water samples were collected from the Second Affiliated Hospital of Dalian Medical University in China.

| 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 186339         | -           | 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 The spot test is used to determine the lytic capacity of the phage. +, clear plaques; -, no plaque; +/-, inconclusive results
c The EOP was determined only for the strains that were positive in the spot test; NT, EOP was not tested; EOP, phage titer on the test strain/phage titer on the host strain; EOP ≥ 0.5 indicates high production efficiency, 0.1 ≤ EOP < 0.5 indicates medium production efficiency, 0.001 ≤ EOP < 0.1 indicates low production efficiency, and EOP < 0.001 indicates inefficient phage production.

**Host range investigation and efficiency-of-plating analysis**

Twenty-three bacterial strains, including *Shigella, Escherichia coli, and Salmonella* (Table 1) were used to determine the lytic capacity of phage Z31 using the spot test method on the basis of its ability to form a lysis zone on lawn cultures of different strains [15]. The purified phage Z31 suspension (10 μl, 10⁹ PFU/ml) was spotted directly onto the surface of a bacterial lawn in a culture plate and incubated overnight at 37°C. The plate was then examined for the appearance of clear zones around the phage drop. Efficiency of plating (EOP) was used to evaluate the host spectrum of the phage by testing a variety of bacterial strains (positive spot test). EOP was calculated as the phage titer on the test strains divided by the phage titer on the host bacteria.
**Transmission electron microscope**

The purified phage suspension (10<sup>9</sup> PFU/ml) was allowed to absorb 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, Tokyo, Japan) [16].

**Phage DNA purification and sequencing**

Phage genomic DNA was extracted from a preparation with a high titer of phage particles (10<sup>10</sup> 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<sup>TM</sup> Nano DNA Sample Prep Kit. Whole-genome sequencing was performed by Shanghai Biozeron Biotechnology Co., Ltd. (Shanghai, China.) using an Illumina NovaSeq 6000 sequencing platform (150 bp × 2) with paired-end reads. A total of 266 Mb of sequence data was obtained. The average read length was 343 bp. Low-quality (Q-value < 20, 97.93%) reads were filtered out using Trimmmomatic v. 0.36, with an approximately 2596x 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 for single-base polymorphisms 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 sequences. 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 in MEGA7 by the neighbor-joining method. A comparative analysis of the complete genome sequences of Z31 and the other members of the genus *Felixounavirus* was performed using EasyFig [21].

**Results and discussion**

A 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 family *Myoviridae*, order *Caudovirales* (Supplementary Fig. S1). A spot test indicated that five *Shigella* strains that were tested and enterotoxigenic *E. coli* K88 CVCC 83902 were lysed by phage Z31 (Table 1). In addition, EOP results revealed that phage Z31 has 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).

Phage Z31 was found to have 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 (Supplementary Table S1). Using tRNAscan-SE, Z31 was found to encode 24 predicted tRNAs (Supplementary Table S2), located between positions 73,864 and 79,210. tRNA genes are universally distributed in dsDNA phages, and virulent phages contain more tRNAs than temperate phages, with higher codon usage bias [22]. It is 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 were predicted to start with an AUG codon (128 ORFs, 96.2%), three with a GUG codon (2.3%), and two with a UUG codon (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%).
A phylogenetic tree constructed using the DNA polymerase sequence (ORF65) revealed that Z31 is 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 formed a cluster that was clearly distinct from those containing members of other genera of the subfamily *Ounaviirinae*, family *Myoviridae* (Fig. 1a). The DNA packaging strategies of tailed dsDNA phages can be classified into six 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. 1b, 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 phages belonging to the genus *Felixounavirus* may use a novel genome packaging strategy that differs from these known strategies.

The genome sequence of phage Z31 was compared to those of the other phages in the branch, using BLASTn and BLASTp. Phage Z31 was found to be similar to enterobacteria phage KhF1 (query coverage, 93%; identity, 96%), enterobacteria phage KhF3 (query coverage, 92%; identity, 95.91%), enterobacteria phage XTG1 (query coverage, 94%; identity, 95.75%), *Escherichia* phage vB_EcoM_LMP25 (query coverage, 91%; identity, 95.97%), and *Escherichia* phage vB_EcoM_AYO145A (query coverage, 90%; identity, 96.52%), and all six of these phages were found to be members of the genus *Felixounavirus* (Fig. 2).

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_EcoMIME338 (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 those of *Salmonella* phage BPS15Q2 (minor fiber protein), *Escherichia* phage vB_EcoM_VpaE1 (tail fiber), *Escherichia coli* phage (tail fiber).
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 evolution 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 conclude that Z31 is a newly isolated phage that can potentially be used as a therapeutic agent.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s00705-021-05121-y.
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.

Declarations

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 KA, Verma NK (2013) Isolation, characterization and comparative genomics of bacteriophage SHV: 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
3. WHO (2014) Antimicrobial resistance: global report on surveillance 2014 Switzerland. WHO Press, Geneva
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. Ann Clin Microbiol Antimicrob 17:1–7
5. Torraca V, Holt K, Mostowy S (2020) Shigella sonnei. Trends Microbiol, pp 1-2 (in press)
6. Warren BR, Parish ME, Schneider KR (2006) Shigella as a foodborne pathogen and current methods for detection in food. Crit Rev Food Technol 46:551–567
7. Reller ME, Nelson JM, Molbak K et al (2006) A large, multiple-restaurant outbreak of infection with Shigella flexneri serotype 2a traced to tomatoes. Clinic Infect Dis 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. Int J Enteric Pathogens 6:41–44
9. Garedev 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. J Infect Public Health 9:348–355
10. Day PT, Nguyen TNT, Thiuy DV et al (2020) Commensal Escherichia coli are a reservoir for the transfer of XDR plasmids into epidemic fluoroquinolone-resistant Shigella sonnei. Nat Microbiol 5:1–9
11. El-Kazzaz SS, Mashaly GE, Zeid MS (2020) Multidrug resistant Shigella ssassociated with class 1 integrase and other virulence genes as a cause of diarrhea in pediatric patients. Open J Med Microbiol 10:1–16
12. Strysko F, Fialkovski V, Marsh Z et al (2019) Outbreak of multidrug-resistant Shigella sonnei infections in a retirement community-Vermont, October-November 2018. Morbid Mortal Wkly Rep 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. Int J Food Microbiol 305:108252
14. Zhang WH, Mi ZQ, Yin XY et al (2013) Characterization of Enterococcus faecalis phage IME-EP1 and its endolysin. PLoS ONE 8:e80436
15. Jun JW, Yun SK, Kim HJ et al (2014) Characterization and complete genome sequence of a novel N4-like bacteriophage, pSb-1 infecting Shigella boydii. Res Microbiol 165:671–678
16. Yuksel S, Thompson K, Ellis A et al (2001) Purification of Piscirickettsia salmonis and associated phage particles. Dis Aquat Org 44:231–235
17. Sambrook J, Russell DW, 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, p 256
18. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:W686–W689
19. Zankari E, Hasman H, Cosentino S et al (2012) Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644
20. Joensen KG, Scheutz F, Lund O et al (2014) Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol 52:1501–1510
21. Sullivan MJ, Petty NK, Beaton SA (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 Res 17:1486–1495
23. Santos SB, Kropinski AM, Ceyssens PJ et al (2011) Genomic and proteomic characterization of the broad-host-range Salmonella phage PVP-SE1: creation of a new phage genus. J Virol 85:11265–11273
24. Casjens SR, Gilcrease EB (2009) Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed bacteriophage virions. Methods Mol Biol 502:91–111
25. Merrill BD, Ward AT, Grosse JH et al (2016) Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. BMC Genom 17:679–694
26. Penadés JR, Chen J, Quiles-Puchalt N et al (2015) Bacteriophage-mediated spread of bacterial virulence genes. Curr Opin Microbiol 23:171–178

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.