



Complete Genome Sequence of vB_EcoP_PR_Kaz2018, a T7-Like Bacteriophage

Madina S. Alexyuk,^a Andrey P. Bogoyavlenskiy,^a Pavel G. Alexyuk,^a Yergali S. Moldakhanov,^a Aizhan S. Turmagambetova,^a Elmira I. Anarkulova,^a Vladimir E. Berezin^a

^aResearch and Production Center for Microbiology and Virology, Laboratory of Antiviral Protection, Almaty, Kazakhstan

ABSTRACT Here, we describe the complete genome sequence of the *Escherichia coli* bacteriophage vB_EcoP_PR_Kaz2018, isolated from a water sample. vB_EcoP_PR_Kaz2018 is a linear double-stranded DNA T7-like podophage with a genome of 39,704 bp containing 45 predicted open reading frames (ORFs).

The *Escherichia coli* bacteriophage vB_EcoP_PR_Kaz2018 was isolated from a water sample and is capable of infecting encapsulated *Escherichia coli* expressing the K1 capsular antigen, a major causative agent of neonatal septicemia, sepsis, and meningitis (1, 2).

Here, we present the complete genome sequence of this bacteriophage, isolated from a freshwater sample collected from the Pravyi Esentai River near the city of Almaty, Kazakhstan. The phage was found to be using as its host a clinical isolate of *Escherichia coli*. The bacterium was cultivated on MacConkey broth at 37°C. The virulence factors associated with the *E. coli* isolate were not investigated in this study. A phage enrichment method was used to isolate the bacteriophage against the host bacterial strain. After incubation, the cocultivation solution was centrifuged at 6,000 rpm, followed by filtration of the supernatant with a 0.45- μ m membrane. The filtrate was tested for phage activity using a spot test and the double-agar overlay method. A single phage plaque was harvested from the agar overlay plate. All of the plaque isolation steps were repeated three times for phage purification.

The genomic DNA was extracted from the phage lysate of the *E. coli* isolate according to the guide for the PureLink viral DNA/RNA minikit (Thermo Fisher Scientific, USA). The DNA library was prepared using the Nextera XT DNA sample preparation kit (Illumina). Whole-genome sequencing was performed with an Illumina MiSeq sequencing platform. Low-quality reads were filtered and adapters trimmed with Trimmomatic (3) from the Genome Detective tool (4). As a result, a total of 3,133,772 paired-end reads were assembled using SPAdes 3.12.0 (5). The average read length after trimming was 286 bp. The genome was sequenced to an average depth of more than 1,000 \times (4, 6). The physical ends of the viral genome were identified by comparing the coverage values of the length of the complete genome of the virus vB_EcoP_PR_Kaz2018 to those of closely related viruses in the NCBI database with BLAST and ViroBlast (7, 8), namely, *Enterobacter* phage K1F (GenBank accession number [AM084414](https://www.ncbi.nlm.nih.gov/nuccore/AM084414)), with 100% query coverage and 95.28% identity; *Escherichia* phage LM33_P1 ([LT594300](https://www.ncbi.nlm.nih.gov/nuccore/LT594300)), with 86% query coverage and 95.22% identity; *Escherichia* phage PE3-1 ([KJ748011](https://www.ncbi.nlm.nih.gov/nuccore/KJ748011)), with 85% query coverage and 95.13% identity; and *Escherichia* phage YZ1 ([MG845865](https://www.ncbi.nlm.nih.gov/nuccore/MG845865)), with 84% query coverage and 94.73% identity.

The complete genome of phage vB_EcoP_PR_Kaz2018 is a linear double-stranded DNA (dsDNA) of 39,704 bp, with 178-bp-long terminal repeats. The GC content of the genome was 49.8%. Gene prediction was done using GeneMark and PHAST (9, 10). A total of 45 putative genes were predicted in the complete genome of the vB_EcoP_

Citation Alexyuk MS, Bogoyavlenskiy AP, Alexyuk PG, Moldakhanov YS, Turmagambetova AS, Anarkulova EI, Berezin VE. 2019. Complete genome sequence of vB_EcoP_PR_Kaz2018, a T7-like bacteriophage. *Microbiol Resour Announc* 8:e01323-19. <https://doi.org/10.1128/MRA.01323-19>.

Editor John J. Dennehy, Queens College

Copyright © 2019 Alexyuk et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Madina S. Alexyuk, Madina.a06@gmail.com, or Andrey P. Bogoyavlenskiy, anpav_63@mail.com.

Received 21 October 2019

Accepted 7 November 2019

Published 5 December 2019

PR_Kaz2018 phage. An analysis of the phage open reading frames (ORFs) using a BLAST search showed that 24 ORFs correspond to proteins with a specific function. The remaining ORFs correspond to hypothetical viral proteins. This genome further increases the known diversity of bacteriophages capable of infecting *E. coli*.

Data availability. The complete genome sequence of the *Escherichia* phage vB_EcoP_PR_Kaz2018 was deposited in GenBank under the accession number [MN510331](https://doi.org/10.1093/nar/gki487). Raw sequence reads are available under BioProject accession number [PRJNA574554](https://doi.org/10.1093/nar/gkr485).

ACKNOWLEDGMENT

This work was supported by grants from the Ministry of Education and Science of Kazakhstan (grants AP05131106 and AP05130916).

REFERENCES

- Polin RA, Harris MC. 2001. Neonatal bacterial meningitis. *Semin Neonatol* 6:157–172. <https://doi.org/10.1053/siny.2001.0045>.
- Simonsen KA, Anderson-Berry AL, Delair SF, Davies HD. 2014. Early-onset neonatal sepsis. *Clin Microbiol Rev* 27:21–47. <https://doi.org/10.1128/CMR.00031-13>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, Pauwels R, Alcantara CL, Eynden EV, Vandamme AM, Deforche K, de Oliveira T. 2019. Genome Detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics* 35:871–873. <https://doi.org/10.1093/bioinformatics/bty695>.
- Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov A, Lesin V, Nikolenko S, Pham S, Prjibelski A, Pyshkin A, Sirotkin A, Vyahhi N, Tesler G, Alekseyev M, Pevzner P. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Deng W, Nickle DC, Learn GH, Maust B, Mullins JI. 2007. ViroBLAST: a stand-alone BLAST Web server for flexible queries of multiple databases and user's datasets. *Bioinformatics* 23:2334–2336. <https://doi.org/10.1093/bioinformatics/btm331>.
- Besemer J, Borodovsky M. 2005. GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* 33:W451–W454. <https://doi.org/10.1093/nar/gki487>.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res* 39:W347–W352. <https://doi.org/10.1093/nar/gkr485>.