Genome Sequences of *Escherichia* Bacteriophages Isolated from Raw Wastewater

Scott P. Keely, Michael P. Herrmann, Asja Korajkic, Nichole E. Brinkman, Brian R. McMinn, G. Shay Fout, Eric N. Villegas

*United States Environmental Protection Agency, National Exposure Research Laboratory Office of Research and Development, Cincinnati, Ohio, USA*

**ABSTRACT**  Somatic coliphages are alternative indicators of fecal pollution and attractive surrogates for viral pathogens. Here, we report the draft genome sequences of three replicate plaques from a novel *Myoviridae* bacteriophage isolated from raw wastewater. These genomes were similar to felix01virus phage and are predicted to contain up to 148 protein-coding genes.

Somatic coliphages are common in fecal matter and can serve as indicators of viral contamination and surrogates for enteric virus removal through wastewater treatment processes (1). However, taxonomic identities of somatic coliphages have not been well described, despite their historical use in plaque assays (2).

Somatic coliphages were isolated from sewage collected at the local wastewater treatment plant using the single agar layer (SAL) method (3). Individual plaques were passaged (three times) using the SAL method, incubated overnight at 37°C with *Escherichia coli* CN-13, centrifuged (15 min, 4,000 rpm), and syringe (0.22 µm) filtered (three times). Free DNA was removed using a TURBO DNA-free kit (Invitrogen, Carlsbad, CA), and nucleic acids were extracted with a PowerViral environmental RNA/DNA isolation kit (Qiagen, Valencia, CA). Genomic DNA was subjected to multiple displacement amplification (MDA) using an illustra GenomiPhi V2 DNA amplification kit (GE Healthcare, Chicago, IL). Libraries were prepared using a Nextera XT DNA library preparation kit (Illumina, San Diego, CA). All procedures were performed according to the manufacturer's instructions. Paired-end sequencing (2 × 250 bp) was performed using a MiSeq reagent kit v2 (Illumina). Genomes were assembled from raw reads (Table 1) with DeBruijn graph SPAdes assembler version 3.10.0 (4) using default error-correction options and resulted in k-mer sizes between 21 and 127. Variation analysis service in PATRIC version 3.5.23 was used to identify single nucleotide polymorphisms and to align the raw reads (5). Annotation was completed with PATRIC (5) and tRNAscan-SE version 2.0 (6). Genome sequences were compared using QUAST version 4.5 (7).

Genomes were named (Table 1) according to the naming and classification guide (8). Forty-seven (LMP25), 1,648 (LMP33), and 4,509 (LMP34) contigs (>500 bp) were generated (Table 1). The largest contigs of LMP25 and LMP33 are complete, whereas the largest contig of LMP34 is a partial genome of the same bacteriophage. Genomes vB_EcoM_LMP33 and vB_EcoM_LMP34 were 99.99% identical to vB_EcoM_LMP25 (two and three nucleotide differences), although vB_EcoM_LMP34 did not include 3,511 bp, presumably due to MDA bias (9). Approximately 62% of the LMP34 raw reads correctly mapped to vB_EcoM_LMP25 (median base coverage, 1,617×), and no nucleotide positions of vB_EcoM_LMP25 had ≤10× coverage, indicating that LMP34 contained enough reads to conclude that it contains the vB_EcoM_LMP25 bacteriophage. QUAST analysis suggested circular permuted genomes. BLAST analysis of vB_EcoM_LMP25 indicated that it was similar (92.9% ± 2.5%; total coverage, 91.4%) to *Escherichia* phage vB_EcoM_AY0145A (GenBank accession number NC_028825) of the felix01virus phage group.

---

Citation Keely SP, Herrmann MP, Korajkic A, Brinkman NE, McMinn BR, Fout GS, Villegas EN. 2019. Genome sequences of *Escherichia* bacteriophages isolated from raw wastewater. *Microbiol Resour Announc* 8:e00135-19. https://doi.org/10.1128/MRA.00135-19.

Editor Catherine Putonti, Loyola University Chicago

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Scott P. Keely, keely.scott@epa.gov, or Asja Korajkic, korajik.asja@epa.gov.

S.P.K. and M.P.H. contributed equally to this work.

Received 4 April 2019
Accepted 3 June 2019
Published 27 June 2019
Bacteriophage vB_EcoM_LMP25 was selected for subsequent analysis because it had the highest N50 value. Annotation identified 20 tRNAs, 2 pseudo-tRNAs genes, and 148 coding sequences. Coding regions predicted 125 hypothetical and 23 known proteins. Phage-associated functions included minor tail protein, tail fiber, baseplate, tail tape measure, large subunit terminase, and lysin. The following amino acids codons were observed: Leu (n = 3), Gln and Lys (n = 2), and Cys, Phe, Gln, Arg, Val, Thr, Gly, Ala, Ile, Asp, Asn, Met, Glu, and Pro (n = 1). The dihydrofolate reductase gene was also identified, but the genome lacked an integrase, suggesting that this bacteriophage may not be capable of lysogeny. The absence of integrase function could be a potential antimicrobial strategy to combat multidrug-resistant bacteria.

**Data availability.** The following bacteriophage genome sequences were submitted to NCBI GenBank: *Escherichia* phage vB_EcoM_LMP25 (accession number MK482688), *Escherichia* phage vB_EcoM_LMP33 (accession number MK482689), and *Escherichia* phage vB_EcoM_LMP34 (accession number MK482690). The BioProject number is PRJNA507261, and the raw reads are available at NCBI SRA (accession numbers SRR8280093, SRR8280094, and SRR8280095).

**ACKNOWLEDGMENTS**

The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to the agency’s administrative review and approved for publication. The views expressed in this article are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

**REFERENCES**

1. McMinn BR, Ashbolt NJ, Korajkic A. 2017. Bacteriophages as indicators of faecal pollution and enteric virus removal. Lett Appl Microbiol 65:11–26. https://doi.org/10.1111/lam.12736.
2. U.S. Environmental Protection Agency. 2015. Review of coliphages as possible indicators of fecal contamination for ambient water quality. 820-R-15-098. U.S. Environmental Protection Agency, Washington, DC.
3. U.S. Environmental Protection Agency. 2001. Method 1602: male-specific (f−) and somatic coliphage in water by single agar layer procedure. EPA 821-R-01. U.S. Environmental Protection Agency, Washington, DC.
4. Bankievich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Aleksyev MA, Pevzner PA. 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.
5. Wattam AR, Davis JJ, Assaf R, Boisvert S, Bretton T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbage JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res 45: D535–D542. https://doi.org/10.1093/nar/gkw1017.
6. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–57. https://doi.org/10.1093/nar/gkw413.
7. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
8. Adriaenssens E, Brister JR. 2017. How to name and classify your phage: an informal guide. Viruses 9:470. https://doi.org/10.3390/v9040470.
9. Brinkman NE, Villegas EN, Garland JL, Keely SP. 2018. Reducing inherent biases introduced during DNA viral metagenome analyses of municipal wastewater. PLoS One 13:e0195350. https://doi.org/10.1371/journal.pone.0195350.

**TABLE 1** GenBank accession numbers and assembly statistics

| Isolate name | Genome name | GenBank accession no. | No. of contigs | No. of reads | N50 value (bp) | Largest contig size (bp) | G+C content (%) | Avg coverage (X) |
|--------------|-------------|-----------------------|---------------|-------------|----------------|------------------------|-----------------|-----------------|
| LMP25        | vB_EcoM_LMP25 | MK482688 | 47 | 1,281,856 | 88,409 | 88,409 | 38.90 | 1,930 |
| LMP33        | vB_EcoM_LMP33 | MK482689 | 1,648 | 892,484 | 737 | 88,409 | 38.90 | 1,072 |
| LMP34        | vB_EcoM_LMP34 | MK482690 | 4,509 | 1,299,808 | 1,218 | 84,898 | 39.01 | 1,054 |

*Calculated for the largest contigs.*