



Complete Genome Sequence of a Jumbo Bacteriophage, *Escherichia* Phage vB_EcoM_EC001

Stevan Cucić,^{a,b} Andrew M. Kropinski,^{c,d} Janet Lin,^b  Cezar M. Khursigara,^b  Hany Anany^{a,b,c}

^aGuelph Research and Development Centre, Agriculture and AgriFood Canada, Guelph, Ontario, Canada

^bDepartment of Molecular and Cellular Biology, College of Biological Science, University of Guelph, Guelph, Ontario, Canada

^cDepartment of Food Science, Ontario Agricultural College, University of Guelph, Guelph, Ontario, Canada

^dDepartment of Pathobiology, Ontario Agricultural College, University of Guelph, Guelph, Ontario, Canada

ABSTRACT Here, we report the genome sequence of a jumbo *Escherichia* phage vB_EcoM_EC001, a myovirus isolated from primary sludge using enterohemorrhagic *Escherichia coli* O157:H7. The genome is 240,200 bp long and has 270 predicted coding sequences, including a tryptophanyl tRNA gene. It belongs to genus *Seoulvirus*.

Enterohemorrhagic *Escherichia coli* (EHEC) can cause severe disease, including hemolytic uremic syndrome (1). Phages infecting EHEC strains are of interest due to their potential utility as biocontrol, therapeutic, and diagnostic agents (2, 3).

Standard approaches to phage isolation are biased toward phages that produce clearly visible plaques. The diversity of coliphages producing clearly visible plaques has been probed extensively. We sought to bias our phage isolation toward large genome size phages by using 0.3% agar instead of the standard 0.4% to 0.7% and selecting pinprick plaques. Pinprick plaques on lower agar concentration overlay can indicate larger phage particles and consequently larger genome sizes (4). Large genome phages encode a preponderance of hypothetical proteins without biochemically characterized homologs and therefore merit investigation (5).

EC001 was isolated in 2017 by enriching the unfiltered supernatant of primary sludge from the Guelph, Ontario, wastewater treatment plant with a cocktail of Shiga toxin-producing *E. coli* (STEC) strains (O157, O111, O26, O103, O145, O121, and O45, with respective strain designations ATCC 700927, HA2018015, HA2018016, HA2018017, HA2018018, HA2018019, and HA2018020), adding 5× of tryptic soy broth (TSB) to the wastewater sample, and incubating the sample for 24 hours at 25°C. Thereafter, a few milliliters of the enriched sample were filtered through 0.45- μ m cellulose acetate syringe filters and spotted on 0.3% agar semisolid TSB (ssTSB) containing *E. coli* O157:H7. A pinprick plaque was passaged three times and propagated using the soft agar method and incubation at 25°C (6).

The PureLink viral DNA/RNA minikit (Thermo Fisher Scientific) was used to extract phage genomic DNA from the crude phage lysate according to the supplier's instructions. DNA elution was done with molecular biology-grade water. DNA was normalized to 200 ng as quantified by fluorimetry (Qubit, Thermo Fisher Scientific) and mechanically sheared using an M220 instrument (Covaris). The next-generation sequencing (NGS) library was prepared using the NxSeq AmpFREE low DNA library kit (Lucigen) sequenced with the MiSeq sequencer (Illumina) using a 2 × 300-bp MiSeq reagent kit v3 (600 cycles) (Illumina) with an average coverage depth of 221-fold. All kits were used according to the manufacturer's protocols. The genome was *de novo* assembled using the SeqManNGen15 (DNASTar) using a base of 200,000 reads (out of ~1.8 million reads) and the approximate genome size observed from pulsed-field gel electrophoresis (~240 Kbp). The draft genome was realigned to that of *Salmonella* phage JN03 (GenBank accession number [MT799840](https://www.ncbi.nlm.nih.gov/nuccore/MT799840)) using progressiveMauve (7). A circular contig of 240,200 bp was obtained. Genome annotation

Editor Kenneth M. Stedman, Portland State University

© Crown copyright 2022. 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 Hany Anany, hany.anany@agr.gc.ca.

The authors declare no conflict of interest.

Received 11 January 2022

Accepted 19 January 2022

Published 3 February 2022

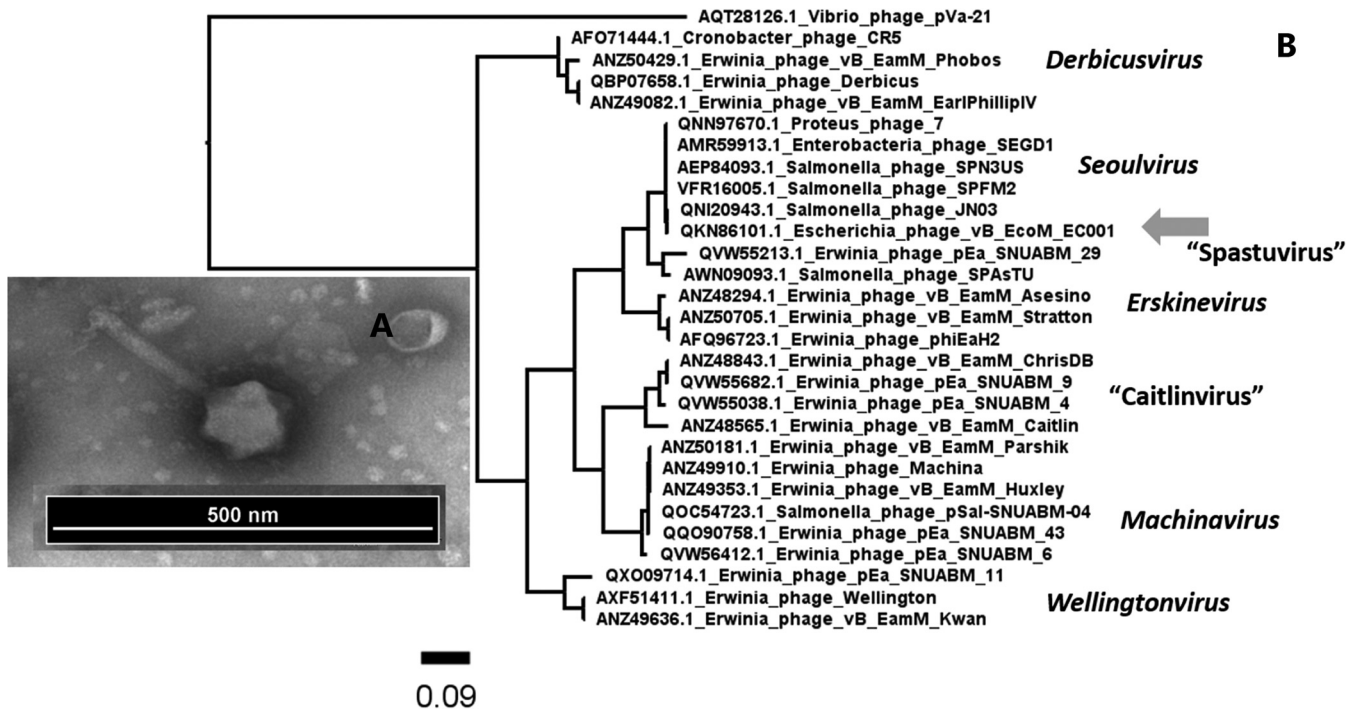


FIG 1 (A) Uranyl acetate transmission electron micrograph of vB_EcoM_EC001 (phage sample was prepared for imaging as in reference 12). (B) Phylogenetic tree generated using phylogeny.fr default settings coupled with FigTree based on the amino acid sequence of large subunit terminase protein. The arrow is pointing to the vB_EcoM_EC001 location in the tree. All selected phages were obtained from ICTV Master Species List 36 (<https://talk.ictvonline.org/files/master-species-lists/m/msl/12314>).

was performed using RAST and MetaGeneAnnotator (8, 9) with protein names verified using Pfam (10). All tools were run with default parameters.

EC001 has icosahedral capsids approximately 120 nm in diameter and tails 190 nm long (Fig. 1A). It has a 240.2-kbp linear, circularly permuted double-stranded DNA (dsDNA) genome with a GC content of 48.5%. Of the 269 predicted protein-coding sequences, only 59 have predicted functions. A phylogenetic tree made using phylogeny.fr (11) based on the amino acid sequence of its terminase large subunit indicates that its closest relative is *Salmonella* phage JN03 (GenBank accession number [MT799840.1](https://www.ncbi.nlm.nih.gov/nuccore/MT799840.1)) (Fig. 1B), a member of the *Seoulvirus* genus.

Data availability. The complete genome sequence of *Escherichia* phage EC001 is accessible at GenBank using the accession number [MN445185.1](https://www.ncbi.nlm.nih.gov/nuccore/MN445185.1). The raw sequence reads are available at NCBI SRA database with accession number [SRP351973](https://www.ncbi.nlm.nih.gov/sra/SRP351973).

ACKNOWLEDGMENTS

This work was funded by Agriculture and Agri-Food Canada through an A-base research grant (grant number 3200) to H.A. and The Natural Sciences and Engineering Research Council of Canada (grant number 371639) to C.M.K.

REFERENCES

- Caprioli A, Morabito S, Brugère H, Oswald E. 2005. Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. *Vet Res* 36:289–311. <https://doi.org/10.1051/vetres:2005002>.
- Hassan AY, Lin JT, Ricker N, Anany H. 2021. The age of phage: friend or foe in the new dawn of therapeutic and biocontrol applications? *Pharmaceuticals* 14:199. <https://doi.org/10.3390/ph14030199>.
- Wan X, Hendrix H, Skurnik M, Lavigne R. 2021. Phage-based target discovery and its exploitation towards novel antibacterial molecules. *Curr Opin Biotechnol* 68:1–7. <https://doi.org/10.1016/j.copbio.2020.08.015>.
- Serwer P, Hayes SJ, Thomas JA, Hardies SC. 2007. Propagating the missing bacteriophages: a large bacteriophage in a new class. *Virology* 361:421–427. <https://doi.org/10.1186/1743-422X-4-21>.
- Yuan Y, Gao M. 2017. Jumbo bacteriophages: an overview. *Front Microbiol* 8:403. <https://doi.org/10.3389/fmicb.2017.00403>.
- Fortier L-C, Moineau S. 2009. Phage production and maintenance of stocks, including expected stock lifetimes, p 203–219. *In* *Bacteriophages*. Springer, New York, NY.
- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The

- RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
9. Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* 15: 387–396. <https://doi.org/10.1093/dnares/dsn027>.
 10. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res* 49: D412–D419. <https://doi.org/10.1093/nar/gkaa913>.
 11. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, Claverie J-M, Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36:W465–W469. <https://doi.org/10.1093/nar/gkn180>.
 12. El-DougDoug NK, Cucic S, Abdelhamid AG, Brovko L, Kropinski AM, Griffiths MW, Anany H. 2019. Control of Salmonella Newport on cherry tomato using a cocktail of lytic bacteriophages. *Int J Food Microbiol* 293:60–71. <https://doi.org/10.1016/j.ijfoodmicro.2019.01.003>.