

Complete Genome Sequence of *Bacillus thuringiensis* Bacteriophage Smudge

Jessica L. Cornell, Eileen Breslin, Zachary Schuhmacher, Madison Himmelright, Cassandra Berluti, Charles Boyd, Rachel Carson, Elle Del Gallo, Caris Giessler, Benjamin Gilliam, Catherine Heatherly, Julius Nevin, Bryan Nguyen, Justin Nguyen, Jocelyn Parada, Blake Sutterfield, Muruj Tukruni, Louise Temple

Department of Integrated Science & Technology, James Madison University, Harrisonburg, Virginia, USA

Smudge, a bacteriophage enriched from soil using *Bacillus thuringiensis* DSM-350 as the host, had its complete genome sequenced. Smudge is a myovirus with a genome consisting of 292 genes and was identified as belonging to the C1 cluster of *Bacillus* phages.

Received 10 May 2016 Accepted 23 June 2016 Published 18 August 2016

Citation Cornell JL, Breslin E, Schuhmacher Z, Himmelright M, Berluti C, Boyd C, Carson R, Del Gallo E, Giessler C, Gilliam B, Heatherly C, Nevin J, Nguyen B, Nguyen J, Parada J, Sutterfield B, Tukruni M, Temple L. 2016. Complete genome sequence of *Bacillus thuringiensis* bacteriophage Smudge. *Genome Announc* 4(4):e00572-16. doi:10.1128/genomeA.00572-16.

Copyright © 2016 Cornell 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 Louise Temple, templelm@jmu.edu.

The bacteriophage Smudge was isolated on *Bacillus thuringiensis* DSM-350 as the host organism from soil samples taken from the James Madison University campus in Harrisonburg, VA, in 2015 (GPS coordinates 38°44'80.81" N, 78°86'59.41" W) by undergraduate researchers in collaboration with the SEA-PHAGES (Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science) program. Because the bacterium *B. thuringiensis* is safe for human use, even though it is commonly used as a biopesticide, it was safe for undergraduates to use in the research lab (1).

Smudge is a myovirus, as shown by the presence of contractile and extended tails. The extended tail was 49.7 ± 2.1 nm, and the contracted tail was 29.5 ± 1.5 nm. The symmetrical head diameter was 23.2 ± 2.4 nm. In this study, the sample soil was enriched using standard procedures to isolate a single phage. Phage genomic DNA was submitted to the Pittsburgh Bacteriophage Institute for Illumina Sequencing. About 50,000 reads of raw data were assembled into a single contig using Newbler (2) with 100-fold average depth of coverage. Gene prediction was completed using GeneMark (3) and Glimmer (4). Smudge was autoannotated using DNA Master (<http://cobamide2.bio.pitt.edu/computer.htm>) and all gene predictions were manually curated. Functions of proteins were predicted using protein BLAST (5), HHPRED, and Conserved Domains Database.

The Smudge genome consisted of 162,040 bp with a 38.8% GC content. Similar to other *Bacillus* phages, the genome contained a direct terminal repeat of 2,820 bp determined by the occurrence of double coverage with well-defined margins in the assembled contig. Smudge is most closely related to Megatron (KJ489401), with 98% query coverage (5). This finding and similarity to other published phages place Smudge in the C1 cluster (6). The number of predicted protein-coding regions found was 292, and no tRNA genes were present, as shown by Aragorn (7). Structural proteins included tail lysins,

baseplate, tail sheath, tail fiber, major capsid, terminase, and prohead protease. Genes involved in DNA replication included DNA helicase, DNA primase, and DNA polymerase. Gene regulatory proteins included sigma factors, two helix-turn-helix binding proteins, and an I-BasI protein. Seven proteins involved in nucleic acid metabolism were adenylate kinase, thymidylate synthase, exonuclease, flavodoxin, thioredoxin, and two ribonucleotide diphosphate reductase beta subunits. The genome also was predicted to contain two endolysins and a holin. There were also two types of recombination proteins, Holliday junction resolvase and recombinase A. Aside from these genes, approximately 1,500 bp in the terminal repeats were noncoding; two hypothetical proteins were predicted in the remaining portion of the terminal repeats.

Accession number(s). The complete genome of bacteriophage Smudge is deposited in GenBank under the accession number [KX156152](https://www.ncbi.nlm.nih.gov/nuclink/KX156152).

ACKNOWLEDGMENTS

We thank teaching assistants Anna Desmarais, Laith Harb, and Laura Lambert. Julius Nevin and Kyle Porter isolated the bacteriophage Smudge.

This work was funded by the Integrated Science and Technology Department and the Honors Program of James Madison University. Howard Hughes Medical Institute initiated and supports the SEA-PHAGES program.

REFERENCES

1. Ibrahim MA, Griko N, Junker M, Bulla LA. 2010. *Bacillus thuringiensis*: a genomics and proteomics perspective. *Bioeng Bugs* 1:31–50. <http://dx.doi.org/10.4161/bbug.1.1.10519>.
2. Liu T, Tsai C, Lee W, Chiang W. 2013. Optimizing information in next-generation-sequencing (NGS) reads for improving *de novo* genome assembly. *PLoS One*. 8:e0069503. <http://dx.doi.org/10.1371/journal.pone.0069503>.
3. Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding

- in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* 33:W451–W454. <http://dx.doi.org/10.1093/nar/gki487>.
4. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23: 673–679. <http://dx.doi.org/10.1093/bioinformatics/btm009>.
 5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. <http://dx.doi.org/10.1006/jmbi.1990.9999>.
 6. Sauder AB, Quinn MR, Brouillette A, Caruso S, Cresawn S, Erill I, Lewis L, Loesser-Casey K, Pate M, Scott C, Stockwell S, Temple L. 2016. Genomic characterization and comparison of seven *Myoviridae* bacteriophage infecting *Bacillus thuringiensis*. *Virology* 489:243–251. <http://dx.doi.org/10.1016/j.virol.2015.12.012>.
 7. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <http://dx.doi.org/10.1093/nar/gkh152>.