Complete Genome Sequence of Cluster C Mycobacteriophage McGee

Imade Y. Nsa,[a] Ayodeji A. Odunsi,[a] Chioma V. Okorie,[a] Seun D. Bamisaye,[a] Rashidat O. Abass,[a] Dennis O. Amadi,[a] Joshua E. Ebitigha,[a] Oluwafemi O. Kadiri,[a] Temidayo M. Makinwa,[a] Solape O. Olaiya,[a] Josephine O. Oluwafemi,[a] Tomilayo L. Owoeye,[a] Zainab A. Sekoni,[a] Abisoye F. Olubajo,[a] Tenny O. G. Egwuatu,[a] Ganiyu O. Oyetibo,[a] Matthew O. Ilori[a,b]

a Department of Microbiology, Faculty of Science, University of Lagos, Lagos, Nigeria
b Institute of Maritime Studies, University of Lagos, Lagos, Nigeria

ABSTRACT Mycobacteriophage McGee is a myovirus isolated from a wet soil sample collected at Manassas, VA, using *Mycobacterium smegmatis* mc²155. McGee has a genome 156,008 bp long, containing 237 protein-coding genes, 31 tRNA genes, and 1 transfer-messenger RNA (tmRNA) gene. McGee shares high gene content similarity to phages in actinobacteriophage cluster C1.

Numerous bacteriophages capable of infecting *Mycobacterium smegmatis* mc² 155 have been isolated and characterized, providing important insights into viral diversity and population structure (1). Here, we report on McGee, a mycobacteriophage newly isolated from soil collected in Manassas, VA (coordinates 38.660592 N, 77.436041 W), using standard procedures (https://seaphagesphagediscoveryguide.helpdocsonline.com/home; accessed 15 July 2022). The soil sample was washed with 7H9 liquid medium, the wash filtered (pore size, 0.02 μm), and the filtrate plated in soft agar overlay containing *Mycobacterium smegmatis* mc² 155. McGee was purified with multiple rounds of plating. Negative-stain transmission electron microscopy revealed McGee to possess a *Myoviridae* morphology (Fig. 1).

DNA was isolated using the Promega Wizard DNA cleanup kit, prepared for sequencing using the NEB Ultra II library kit (New England Biolabs, Ipswich, MA), and sequenced on an Illumina MiSeq instrument (v3 reagents), resulting in 77,248 single-end 150-bp reads providing 73× coverage. The raw reads were assembled and assessed using the programs Newbler v2.9 (2) and Consed v23 (3), respectively. The assembly yielded a circularly permuted genome of 156,008 bp with 64.6% GC content.

Genes were identified and annotated using DNA Master v5.23.6 (http://cobamide2.bio.pitt.edu/) (4), GLIMMER v3.02 (5), GeneMark v2.5 (6), and the Starterator v463 database (http://phages.wustl.edu/starterator), whereas gene functional assignments were determined using BLASTp searches against the NCBI nonredundant v2.13.0 (7) and actinobacteriophage databases (8), HHpred (9), and Phamerator (10). Aragorn v1.2.41 (11) and tRNAscan v2.0.9 (12) were used to call tRNAs and transfer-messenger RNAs (tmRNAs). Membrane proteins were identified using SOSUI v1.1 (13–15) and TMHMM v2.0 (16, 17). All software was run using default parameters.

A total of 269 putative genes were identified in McGee, including 237 protein-coding genes, 31 tRNAs, and 1 tmRNA. Among those that could be assigned a function, the virion structure and assembly genes include a major capsid protein (gp101), capsid decoration protein (gp103, gp117), tail sheath protein (gp129), tail assembly chaperones (gp132, gp133), tape measure protein (gp134), minor tail proteins (gp138, gp146, gp147), and baseplate wedge protein (gp141, gp142). The genes involved in McGee's DNA metabolism include DNA helicase (gp188), thymidylate kinase (gp197), DnaC-like helicase loader (gp201), DnaB-like double-stranded DNA (dsDNA) helicase (gp202),
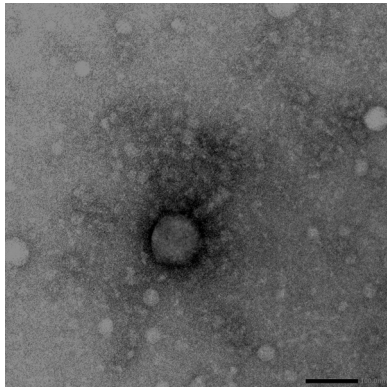
FIG 1 Transmission electron micrograph of McGee. Phage lysate was mounted on a Formvar copper grid and stained with 1% uranyl acetate. McGee has a *Myoviridae* morphotype, with a capsid 103 nm in diameter and a tail 86 nm long ($n = 1$).

DNA primase (gp204), DnaJ-like chaperonin (gp206), single-stranded DNA (ssDNA) binding protein (gp207), DnaE-like DNA polymerase III α (gp208), RF-1 peptide chain release factor (gp209), RecA-like DNA recombinase (gp210), Holliday junction resolvase (gp210), RusA-like resolvase (gp214), Ro-like RNA binding protein (gp233), and multiple proteins with a helix-turn-helix DNA binding domain (gp54, gp61, gp62, gp196). Genes involved in lysis, lysin A, holin, and lysin B are encoded by gp252, gp253, and gp254, respectively. The largest gene in McGee is a hypothetical protein gp98 (4,113 bp long) and the smallest gp64 (75 bp long).

Based on its gene content similarity (GCS) of at least 35% to phages in the Actinobacteriophage database (https://phagesdb.org/), McGee was assigned to phage cluster C1 (8, 18, 19). McGee shares the highest GCS with Kboogie (95.98%) and encodes 3 additional HNH endonucleases (gp148, gp221, gp251) relative to Kboogie.

Data availability. The complete genome sequence for McGee can be found at GenBank under accession number ON637764, while its raw reads have been deposited under Sequence Read Archive accession number SRX14483217.

ACKNOWLEDGMENTS

REFERENCES

1. Hatfull GF. 2018. Mycobacteriophages. Microbiol Spectr 6:GPP3-0026-2018. https://doi.org/10.1128/microbiolspec.GPP3-0026-2018.
2. Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next generation sequencing data. Genomics 95:315–327. https://doi.org/10.1016/j.ygeno.2010.03.001.
3. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res 8:195–202. https://doi.org/10.1101/gr.8.3.195.
4. Pope WH, Jacobs-Sera D. 2018. Annotation of bacteriophage genome sequences using DNA master: an overview. Methods Mol Biol 1681:217–229. https://doi.org/10.1007/978-1-4939-7343-9_16.
5. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679. https://doi.org/10.1093/bioinformatics/btm009.
6. 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.
7. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
8. Russell DA, Hatfull GF. 2017. PhagesDB: the actinobacteriophage database. Bioinformatics 33:784–786. https://doi.org/10.1093/bioinformatics/btw711.
9. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248. https://doi.org/10.1093/nar/gki408.
10. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12:395. https://doi.org/10.1186/1471-2105-12-395.
11. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.

12. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–964. https://doi.org/10.1093/nar/25.5.955.

13. Hirokawa T, Boon-Chieng S, Mitaku S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14:378–379. https://doi.org/10.1093/bioinformatics/14.4.378.

14. Mitaku S, Hirokawa T. 1999. Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. Protein Eng 12:953–957. https://doi.org/10.1093/protein/12.11.953.

15. Mitaku S, Hirokawa T, Tsuji T. 2002. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. Bioinformatics 18:608–616. https://doi.org/10.1093/bioinformatics/18.4.608.

16. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.

17. Sonnhammer ELL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences, p 175–182. In Glasgow J, Littlejohn T, Major F, Lathrop R, Sankoff D, Sensen C (ed), Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, CA.

18. Hatfull GF. 2012. The secret lives of mycobacteriophages. Adv Virus Res 82:179–288. https://doi.org/10.1016/b978-0-12-394621-8.00015-7.

19. Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, Russell DA, Warner MH, Hatfull GF, Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES). 2017. Bacteriophages of Gordonia spp. display a spectrum of diversity and genetic relationships. mBio 8:e01069-17. https://doi.org/10.1128/mBio.01069-17.