Complete Genome Sequence of Enterotoxigenic *Escherichia coli* Myophage Murica

Joseph N. Wilder, Jacob C. Lancaster, Jesse L. Cahill, Eric S. Rasche, Gabriel F. Kuty Everett

Center for Phage Technology, Texas A&M University, College Station, Texas, USA

Murica is an rv5-like myophage that infects enterotoxigenic *Escherichia coli*. Pathogenic *E. coli* strains are responsible for many intestinal diseases, and phages that infect these bacteria may prove useful in preventing severe health issues. The following is a report of the complete genome sequence of Murica and its important features.

Address correspondence to Gabriel F. Kuty Everett, cpt@tamu.edu.

Enterotoxigenic *Escherichia coli* strains serve as one of the leading causes of moderate to severe diarrhea in developing countries, affecting more people than *Shigella* or *Salmonella* species in many areas (1–3). The U.S. Food and Drug Administration has reported that bacteriophages are safe for use as a food additive in some meat and poultry products, paving the way for the production and use of phage-based products against enterotoxigenic *E. coli* (4). Here, we describe the complete genome sequence of a novel rv5-like myophage, Murica, which is active against enterotoxigenic *E. coli*.

Bacteriophage Murica was isolated from a pig fecal sample collected at College Station, TX. Phage DNA was sequenced in an Illumina MiSeq 250-bp paired-end run with a 550-bp insert library at the Genomic Sequencing and Analysis Facility at the University of Texas (Austin, TX). Quality controlled trimmed reads were assembled to a single contig of circular assembly at 86.6-fold coverage using SPAdes version 3.5.0 (5). The contig was confirmed to be complete by PCR using primers that face the upstream and downstream ends of the contig. Products from the PCR amplification of the junctions of concatemeric molecules were sequenced by Sanger sequencing (Eton Bioscience, San Diego, CA). Genes were predicted using GeneMarkS (6) and corrected using software tools available on the Center for Phage Technology (CPT) Galaxy instance (https://cpt.tamu.edu/galaxy-public/). Morphology was determined using transmission electron microscopy performed at the Texas A&M University Microscopy and Imaging Center.

Murica contains a 135,391-bp genome that has a coding density of 90.7%. It has a G+C content of 43.6%, which is lower than the average G+C content for enterotoxigenic *E. coli* strains (50%) (7). The genome contains 212 putative coding sequences, 49 of which code for predicted functions based on InterProScan and BLASTp analysis (8, 9). Murica shares 91.5% nucleotide sequence identity across the genome with *E. coli* myophage rv5, as determined by Emboss Stretcher (10). It is a member of the Lytic22 cluster recently described by Grose and Casjens (11). Murica was opened to the *rIIa* gene in accordance with the precedent set by rv5. Seven tRNA genes were identified compared to the five found in rv5 (12).

Murica has a 2,554-bp region that is comparable to the 2,604-bp noncoding region in rv5; however, the gap in Murica appears to contain a novel gene of unknown function. Rv5 has an HNH homing endonuclease interrupting the large terminase, while the large terminase remains intact in Murica. The tail fiber protein (Murica36) shows 99% amino acid identity with the tail fiber protein of *E. coli* O157:H7 phage vB_EcoM_FFH2 (accession no. NC_024134), a phage that has been described as having broad host range and was recently shown to be effective in reducing *E. coli* O157:H7 contamination in spinach and beef (13, 14). This suggests that Murica may also have a broad host range and may be useful in the treatment of contaminated food products.

Nucleotide sequence accession number. The genome sequence of phage Murica was contributed as accession no. KT001917 to GenBank.

ACKNOWLEDGMENTS

REFERENCES

1. **Patzi-Vargas S, Zaidi MB, Perez-Martinez I, León-Cen M, Michel-Ayala A, Chaussabel D, Estrada-Garcia T**. 2015. Diarrheagenic *Escherichia coli* carrying supplementary virulence genes are an important cause of moderate to severe diarrheal disease in Mexico. PLoS Negl Trop Dis 9:e0003510. http://dx.doi.org/10.1371/journal.pntd.0003510.
2. **Lozer DM, Souza TB, Monfardini MV, Vicentini F, Kitagawa SS, Scaletsky IC, Spano LC**. 2013. Genotypic and phenotypic analysis of diarrheagenic *Escherichia coli* strains isolated from Brazilian children living in low socioeconomic level communities. BMC Infect Dis 13:418. http://dx.doi.org/10.1186/1471-2334-13-418.
3. **Moyo SJ, Maselle SY, Matee MI, Langeland N, Mylvaganam H**. 2007. Identification of diarrheagenic *Escherichia coli* isolated from infants and children in Dar es Salaam, Tanzania. BMC Infect Dis 7:92. http://dx.doi.org/10.1186/1471-2334-7-92.

4. **Atterbury RJ**. 2009. Bacteriophage biocontrol in animals and meat products. Microb Biotechnol **2**:601–612. http://dx.doi.org/10.1111/j.1751-7915.2009.00089.x.

5. **Bankevich 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, Alekseyev MA, Pevzner PA**. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19**:455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

6. **Besemer J, Lomsadze A, Borodovsky M**. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res **29**:2607–2618. http://dx.doi.org/10.1093/nar/29.12.2607.

7. **Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, Cunningham AF, Petty NK, Mahon V, Brinkley C, Hobman JL, Savarino SJ, Turner SM, Pallen MJ, Penn CW, Parkhill J, Turner AK, Johnson TJ, Thomson NR, Smith SG, Henderson IR**. 2010. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407. J Bacteriol **192**:5822–5831. http://dx.doi.org/10.1128/JB.00710-10.

8. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL**. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**:421. http://dx.doi.org/10.1186/1471-2105-10-421.

9. **Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C**. 2009. InterPro: the integrative protein signature database. Nucleic Acids Res **37**:D211–D215. http://dx.doi.org/10.1093/nar/gkn785.

10. **Myers EW, Miller W**. 1988. Optimal alignments in linear space. Comput Appl Biosci **4**:11–17. http://dx.doi.org/10.1093/bioinformatics/4.1.11.

11. **Grose JH, Casjens SR**. 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology **468–470C**:421–443.

12. **Kropinski AM, Waddell T, Meng J, Franklin K, Ackermann HW, Ahmed R, Mazzocco A, Yates J III, Lingohr EJ, Johnson RP**. 2013. The host-range, genomics and proteomics of Escherichia coli O157:H7 bacteriophage rV5. Virol J **10**:76. http://dx.doi.org/10.1186/1743-422X-10-76.

13. **Hong Y, Pan Y, Harman NJ, Ebner PD**. 2014. Complete genome sequences of two Escherichia coli O157:H7 phages effective in limiting contamination of food products. Genome Announc **2**(5):e00519-14. http://dx.doi.org/10.1128/genomeA.00519-14.

14. **Hong Y, Pan Y, Ebner PD**. 2014. Meat Science and Muscle Biology Symposium: development of bacteriophage treatments to reduce Escherichia coli O157:H7 contamination of beef products and produce. J Anim Sci **92**:1366–1377. http://dx.doi.org/10.2527/jas.2013-7272.