Complete Genome Sequence of *Salmonella enterica* Serovar Heidelberg Myophage Meda

John K. Neff,[a] Yicheng Xie,[a] Jason J. Gill,[a] Mei Liu[a]

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

ABSTRACT *Salmonella enterica* serovar Heidelberg is a multidrug-resistant food-borne pathogen that originated from poultry and cattle. Bacteriophages isolated for this pathogen may be used as biocontrol agents in food products or animals for preventing *Salmonella* foodborne diseases. Here, we present the complete genome sequence of *Salmonella* Heidelberg phage Meda.

*S*almonella enterica* serovar Heidelberg is a multidrug-resistant foodborne pathogen that originated from poultry and cattle (1). As antibiotic resistance increases in foodborne pathogens, phages may be used as biocontrol agents in food or animal production to combat this rising threat.

Myophage Meda was isolated from the soil of a cattle-harvesting facility in Michigan in August 2016 by growing *Salmonella* Heidelberg in the soil extract using tryptic soy broth (Difco) at 37°C with aeration. Phages were isolated and propagated by the soft agar overlay method (2). Phage genomic DNA was prepared using a modified Promega Wizard DNA cleanup kit protocol, as described previously (3). Pooled indexed DNA libraries were prepared using the Illumina TruSeq Nano low-throughput (LT) kit, and the sequence was obtained from the Illumina MiSeq platform using the MiSeq V2 500-cycle reagent kit, following the manufacturer's instructions, producing 312,918 paired-end reads for the index containing the phage genome. Reads were quality controlled in FastQC 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), trimmed with the FASTX-Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/), and assembled using SPAdes 3.5.0 (4). Meda was assembled at 223.7-fold coverage. The assembled contig completion was confirmed by PCR using primers (5'-TGAGCATGGTTTCCGTTAG AG-3' and 5'-GTGATTCTAGGCCAGTTGGTAG-3') facing off the ends of the assembled contig and Sanger sequencing of the resulting product, with the contig sequence manually corrected to match the resulting Sanger sequencing read. GLIMMER 3.0 (5) and MetaGeneAnnotator 1.0 (6) were used to predict protein-coding genes, along with manual curation. tRNA genes were predicted using ARAGORN 2.36 (7). Protein functions were predicted using BLASTp 2.2.28 (8) to detect sequence homology. Protein structure prediction was performed using HHpred 2.1 (9), and conserved domain searches were conducted using InterProScan 5.15-5.40 (10). All analyses were done with default settings using the CPT Galaxy (11) (cpt.tamu.edu) and Web Apollo (12) interfaces.

The complete Meda genome is 84,668 bp long, consisting of 131 protein-coding genes with a coding density of 88.04%. It has a GC content of 38.85%, which is significantly lower than that of its host, *Salmonella* Heidelberg (52.08%) (13). Meda shares the most protein homology with *Salmonella* bacteriophage Felix O1 (14), with 130 similar proteins at an E value of <0.001 by BLASTp, using the GenBank non-redundant (nr) database. Like Felix O1, Meda appears not to possess a copy of RNA polymerase or any proteins homologous to known host RNA polymerase modifiers. Meda's lysis genes are distributed across the genome rather than forming a discrete lysis cassette. While spanins and an endolysin (glycoside hydrolase) were annotated, a

holin was not identified. The Meda genome contains several HNH endonucleases, one of which interrupts the large subunit of anaerobic nucleoside diphosphate reductase. Another intron containing no HNH endonuclease sequence is present in a DNA polymerase gene. A predicted polynucleotide kinase gene was identified in Meda, but the protein-coding sequence is split into two reading frames (GenBank accession numbers AXY86388 and AXY86389) by an intervening stop codon. The presence of this stop codon was confirmed by Sanger sequencing. It is not clear if this gene is essential for phage growth or if it is still functional despite this interruption.

Data availability. The genome sequence of phage Meda has been deposited under GenBank accession number MH586731. The associated BioProject, SRA, and Bio-Sample accession numbers are PRJNA222858, SRR8787572, and SAMN11259651, respectively.

REFERENCES

1. Taylor AL, Murphree R, Ingram LA, Garman K, Solomon D, Coffey E, Walker D, Rogers M, Marder E, Bottomley M, Woron A, Thomas L, Roberts S, Hardin H, Arjmandi P, Green A, Simmons L, Cornell A, Dunn J. 2015. Multidrug-resistant Salmonella Heidelberg associated with mechanically separated chicken at a correctional facility. Foodborne Pathog Dis 12: 950–952. https://doi.org/10.1089/fpd.2015.2008.
2. Adams MK. 1959. Bacteriophages. Interscience Publishers, Inc., New York, NY.
3. Gill JJ, Berry JD, Russell WK, Lessor L, Escobar-Garcia DA, Hernandez D, Kane A, Keene J, Maddox M, Martin R, Mohan S, Thorn AM, Russell DH, Young R. 2012. The Caulobacter crescentus phage phiCbK: genomics of a canonical phage. BMC Genomics 13:542. https://doi.org/10.1186/1471-2164-13-542.
4. 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. https://doi.org/10.1089/cmb.2012.0021.
5. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27: 4636–4641. https://doi.org/10.1093/nar/27.23.4636.
6. 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.
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. https://doi.org/10.1093/nar/gkh152.
8. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
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. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
11. Cock PJ, Gruning BA, Paszkiewicz K, Pritchard L. 2013. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. PeerJ 1:e167. https://doi.org/10.7717/peerj.167.
12. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE. 2013. Web Apollo: a Web-based genomic annotation editing platform. Genome Biol 14:R93. https://doi.org/10.1186/gb-2013-14-8-r93.
13. Hoffmann M, Luo Y, Lafon PC, Timme R, Allard MW, McDermott PF, Brown EW, Zhao S. 2013. Genome sequences of Salmonella enterica serovar Heidelberg isolates isolated in the United States from a multistate outbreak of human Salmonella infections. Genome Announc 1:e00004-12. https://doi.org/10.1128/genomeA.00004-12.
14. Whichard JM, Weigt LA, Borris DJ, Li LL, Zhang Q, Kapur V, Pierson FW, Lingohr EJ, She YM, Kropinski AM, Sriranganathan N. 2010. Complete genomic sequence of bacteriophage Felix O1. Viruses 2:710–730. https://doi.org/10.3390/v2030710.