



# Complete Genome Sequence of *Salmonella enterica* Siphophage Shemara

Michael Chung,<sup>a</sup> Yicheng Xie,<sup>a</sup> Heather Newkirk,<sup>a</sup> Mei Liu,<sup>a</sup>  Jason J. Gill,<sup>a</sup>  Jolene Ramsey<sup>a</sup>

<sup>a</sup>Center for Phage Technology, Texas A&M University, College Station, Texas, USA

**ABSTRACT** Here, we present the annotated genome of Shemara, a siphophage of *Salmonella enterica*. The Shemara genome is 44 kb with 83 predicted protein-coding genes. At the nucleotide and amino acid levels, Shemara is most similar to phages in the *Guernseyvirinae* subfamily.

Here, we describe the genome of phage Shemara, a siphophage that infects the Gram-negative bacterium *Salmonella enterica*. Humans are typically exposed to *S. enterica*, one of the most common causes of foodborne illnesses, through infected livestock or plant food products (1). Significant progress has been made in the investigation and application of phages as natural biocontrol agents in food products to reduce the incidence of disease (2).

Bacteriophage Shemara was sourced from a south Texas beef cattle feedlot. Shemara was isolated against *S. enterica* serotype Anatum, which is associated with recent human disease outbreaks (3, 4). Host bacteria were cultured on tryptic soy broth or agar (Difco) at 37°C with aeration, and phage isolation and propagation were done using the soft-agar overlay method (5). The phage morphology was viewed after negative staining with 2% (wt/vol) uranyl acetate under transmission electron microscopy at the Texas A&M Microscopy and Imaging Center (6). Genomic DNA was purified by the shotgun assembly modification of the Wizard DNA kit (Promega) described by Summer (7). After purification and library preparation using an Illumina TruSeq Nano low-throughput kit, the phage DNA was sequenced on the Illumina MiSeq platform using paired-end 250-bp reads. The 128,596 total resulting sequence reads were quality controlled with FastQC ([www.bioinformatics.babraham.ac.uk/projects/fastqc](http://www.bioinformatics.babraham.ac.uk/projects/fastqc)) and trimmed with the FASTX Toolkit v0.0.14 ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). A single contig was assembled with SPAdes v3.5.0 to 130-fold coverage at default parameters, including k-mers of 21, 33, and 55 (8). A second contig, assembled using k-mers of 99, 103, and 127, with 36.7-fold coverage, was opened in a different place, allowing complete closure of the genome sequence. Despite this, PhageTerm analysis with raw sequencing reads was unable to assign the terminus type (9). Identification of gene sequences was done with MetaGeneAnnotator v1.0, GLIMMER v3.0, and ARAGORN v2.36 (10–12). TransTermHP v2.09 was used to predict the presence of rho-independent terminators (13). Gene function was predicted using conserved domain searches with InterProScan v5.22-61 and BLAST similarity searches with a 0.001 maximum expectation value threshold against the NCBI nonredundant and UniProtKB Swiss-Prot and TrEMBL databases (14–16). TMHMM v2.0 predicted transmembrane domains (17). Annotation tools were used in the Galaxy and Web Apollo instances hosted by the Center for Phage Technology (<https://cpt.tamu.edu/galaxy-pub/>) (18, 19) with default parameters, unless otherwise stated. ProgressiveMauve v2.4.0 assessed the whole-genome nucleotide similarity (20).

Shemara is a siphophage, and its genomic contig has a length of 44,342 bp with 51.4% G+C content. With 83 predicted gene-coding sequences, 34 of which were

**Citation** Chung M, Xie Y, Newkirk H, Liu M, Gill JJ, Ramsey J. 2020. Complete genome sequence of *Salmonella enterica* siphophage Shemara. Microbiol Resour Announc 9:e01518-19. <https://doi.org/10.1128/MRA.01518-19>.

**Editor** John J. Dennehy, Queens College

**Copyright** © 2020 Chung 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 Jolene Ramsey, [jolener@tamu.edu](mailto:jolener@tamu.edu).

**Received** 11 December 2019

**Accepted** 13 January 2020

**Published** 6 February 2020

assigned a putative function, its coding density is 95.8%. Shemara is most similar to several phages within the *Guernseyvirinae* subfamily (21), namely, *Salmonella* virus VSip (GenBank accession number [MH424444](#)), with 72% nucleotide identity and 59 shared proteins, and *Salmonella* phage St162 (GenBank accession number [MF158037](#)), with 72.1% nucleotide identity and 53 shared proteins.

**Data availability.** The genome sequence and associated data for phage Shemara were deposited under GenBank accession number [MN070121](#), BioProject accession number [PRJNA222858](#), SRA accession number [SRR8869235](#), and BioSample accession number [SAMN11360359](#).

## ACKNOWLEDGMENTS

This work was supported by funding from the National Science Foundation (awards EF-0949351 and DBI-1565146) and from the National Cattlemen's Beef Association and Texas Beef Cattle. Additional support came from the Center for Phage Technology (CPT), an Initial University Multidisciplinary Research Initiative supported by Texas A&M University and Texas A&M AgriLife, and from the Texas A&M University Department of Biochemistry and Biophysics.

We are grateful for the advice and support of the CPT staff.

This announcement was prepared in partial fulfillment of the requirements for BICH464 Bacteriophage Genomics, an undergraduate course at Texas A&M University.

## REFERENCES

- Nair DVT, Venkitanarayanan K, Kollanoor Johny A. 2018. Antibiotic-resistant *Salmonella* in the food supply and the potential role of antibiotic alternatives for control. *Foods* 7:167. <https://doi.org/10.3390/foods7100167>.
- Oh J-H, Park M-K. 2017. Recent trends in *Salmonella* outbreaks and emerging technology for biocontrol of *Salmonella* using phages in foods: a review. *J Microbiol Biotechnol* 27:2075–2088. <https://doi.org/10.4014/jmb.1710.10049>.
- Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ. 2013. Outbreak-associated *Salmonella enterica* serotypes and food commodities, United States, 1998–2008. *Emerg Infect Dis* 19:1239–1244. <https://doi.org/10.3201/eid1908.121511>.
- Xie Y, Savell JW, Arnold AN, Gehring KB, Gill JJ, Taylor TM. 2016. Prevalence and characterization of *Salmonella enterica* and *Salmonella* bacteriophages recovered from beef cattle feedlots in south Texas. *J Food Prot* 79:1332–1340. <https://doi.org/10.4315/0362-028X.JFP-15-526>.
- Adams MH. 1956. *Bacteriophages*. Interscience Publishers, Inc., New York.
- Valentine RC, Shapiro BM, Stadtman ER. 1968. Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from *Escherichia coli*. *Biochemistry* 7:2143–2152. <https://doi.org/10.1021/bi00846a017>.
- Summer EJ. 2009. Preparation of a phage DNA fragment library for whole genome shotgun sequencing. *Methods Mol Biol* 502:27–46. [https://doi.org/10.1007/978-1-60327-565-1\\_4](https://doi.org/10.1007/978-1-60327-565-1_4).
- Bankovich 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>.
- Garneau JR, Depardieu F, Fortier L-C, Bikard D, Monot M. 2017. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Sci Rep* 7:8292. <https://doi.org/10.1038/s41598-017-07910-5>.
- 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>.
- 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>.
- 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>.
- Kingsford CL, Ayanbule K, Salzberg SL. 2007. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 8:R22. <https://doi.org/10.1186/gb-2007-8-2-r22>.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- 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>.
- UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–D515. <https://doi.org/10.1093/nar/gky1049>.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 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>.
- Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltmann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46:W537–W544. <https://doi.org/10.1093/nar/gky379>.
- Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elisk 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>.
- 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>.
- Segovia C, Vasquez I, Maracaja-Coutinho V, Robeson J, Santander J. 2015. Complete genome sequence of *Salmonella enterica* serovar Enteritidis bacteriophage f18SE, isolated in Chile. *Genome Announc* 3:e00600-15. <https://doi.org/10.1128/genomeA.00600-15>.