Complete Genome Sequence of *Salmonella enterica* Serovar Typhimurium Siphophage Seabear

Ketki Patil,a Chi Zeng,b Chandler O’Leary,a Lauren Lessor,a Rohit Kongari,a Jason Gill,a Mei Liu,a

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

bSchool of Biology and Pharmaceutical Engineering, Wuhan Polytechnic University, Wuhan, China

**ABSTRACT** *Salmonella enterica* serovar Typhimurium is a foodborne pathogen that causes gastroenteritis. Due to increases in antibiotic resistance, bacteriophage therapy may be an alternative method for preventing *Salmonella* foodborne infections. We report here the complete genome sequence of a T5-like phage, Seabear, which was isolated against *S. Typhimurium*.

*Salmonella enterica* serovar Typhimurium is a Gram-negative bacterium known to cause acute gastroenteritis leading to diarrhea, vomiting, fever, and abdominal cramps (1, 2). Increasing drug resistance found among strains of *S. Typhimurium* has made phage therapy a control measure for *S. Typhimurium* (3, 4). In this paper, we report the genomic characterization of a T5-like siphophage that infects *S. Typhimurium*.

Phage Seabear was isolated from a wastewater treatment plant in College Station, TX, in 2015 using a poultry isolate of *S. Typhimurium* as the host. Host bacteria were cultured on tryptic soy broth or agar (Difco) at 37°C with aeration. Phage were cultured and propagated using the soft-agar overlay method (5). Seabear was identified as a siphophage using negative-stain transmission electron microscopy performed at the Texas A&M University Microscopy and Imaging Center, as described previously (6). Phage genomic DNA was prepared using a modified Promega Wizard DNA cleanup kit protocol, as described previously (6). Pooled indexed DNA libraries were prepared using the Illumina TruSeq Nano low-throughput (LT) kit. The sequence was obtained from the Illumina MiSeq platform using the MiSeq V2 500-cycle reagent kit, following the manufacturer’s instructions, producing 477,731 paired-end reads for the index containing the phage Seabear genome. FastQC 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to quality control the reads. The reads were trimmed with FASTX-Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/download.html) before being assembled using SPAdes 3.5.0 (7). Contig completion was confirmed by PCR using primers (5’-AAACCATACCCGCAACCAC-3’ and 5’-CATGAGAAAAACCGCTGTA-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 (8) and MetaGeneAnnotator 1.0 (9) were used to predict protein-coding genes with manual verification, and tRNA genes were predicted with ARAGORN 2.36 (10). Rho-independent termination sites were identified via TransTerm (http://transterm.cbcbachem.umd.edu/). Sequence similarity searches were done by BLASTp 2.2.28 (11) against the NCBI nr, UniProt Swiss-Prot (12), and TrEMBL databases. InterProScan 5.15-54.0 (13), LipoP (14), and TMHMM v2.0 (15) were used to predict protein function.

Seabear was assembled at 57.7-fold coverage into a 112,472-bp genome with an 8,522-bp direct terminal repeat, as determined by PhageTerm (18). Seabear has 40.4% GC content, which is lower than that of its host (52.2%) (19). A total of 178 protein-coding genes were predicted with manual verification, and tRNA genes were predicted with ARAGORN 2.36 (10). Rho-independent termination sites were identified via TransTerm (http://transterm.cbcbachem.umd.edu/). Sequence similarity searches were done by BLASTp 2.2.28 (11) against the NCBI nr, UniProt Swiss-Prot (12), and TrEMBL databases. InterProScan 5.15-54.0 (13), LipoP (14), and TMHMM v2.0 (15) were used to predict protein function. All analyses were conducted at default settings via the CPT Galaxy (16) and Web Apollo (17) interfaces (https://cpt.tamu.edu/galaxy-pub).

Seabear was assembled at 57.7-fold coverage into a 112,472-bp genome with an 8,522-bp direct terminal repeat, as determined by PhageTerm (18). Seabear has 40.4% GC content, which is lower than that of its host (52.2%) (19). A total of 178 protein-coding genes were predicted with manual verification, and tRNA genes were predicted with ARAGORN 2.36 (10). Rho-independent termination sites were identified via TransTerm (http://transterm.cbcbachem.umd.edu/). Sequence similarity searches were done by BLASTp 2.2.28 (11) against the NCBI nr, UniProt Swiss-Prot (12), and TrEMBL databases. InterProScan 5.15-54.0 (13), LipoP (14), and TMHMM v2.0 (15) were used to predict protein function. All analyses were conducted at default settings via the CPT Galaxy (16) and Web Apollo (17) interfaces (https://cpt.tamu.edu/galaxy-pub).

Seabear was assembled at 57.7-fold coverage into a 112,472-bp genome with an 8,522-bp direct terminal repeat, as determined by PhageTerm (18). Seabear has 40.4% GC content, which is lower than that of its host (52.2%) (19). A total of 178 protein-coding genes were predicted with manual verification, and tRNA genes were predicted with ARAGORN 2.36 (10). Rho-independent termination sites were identified via TransTerm (http://transterm.cbcbachem.umd.edu/). Sequence similarity searches were done by BLASTp 2.2.28 (11) against the NCBI nr, UniProt Swiss-Prot (12), and TrEMBL databases. InterProScan 5.15-54.0 (13), LipoP (14), and TMHMM v2.0 (15) were used to predict protein function. All analyses were conducted at default settings via the CPT Galaxy (16) and Web Apollo (17) interfaces (https://cpt.tamu.edu/galaxy-pub).

Seabear was assembled at 57.7-fold coverage into a 112,472-bp genome with an 8,522-bp direct terminal repeat, as determined by PhageTerm (18). Seabear has 40.4% GC content, which is lower than that of its host (52.2%) (19). A total of 178 protein-coding genes were predicted with manual verification, and tRNA genes were predicted with ARAGORN 2.36 (10). Rho-independent termination sites were identified via TransTerm (http://transterm.cbcbachem.umd.edu/). Sequence similarity searches were done by BLASTp 2.2.28 (11) against the NCBI nr, UniProt Swiss-Prot (12), and TrEMBL databases. InterProScan 5.15-54.0 (13), LipoP (14), and TMHMM v2.0 (15) were used to predict protein function. All analyses were conducted at default settings via the CPT Galaxy (16) and Web Apollo (17) interfaces (https://cpt.tamu.edu/galaxy-pub).
coding sequences and 29 tRNAs were annotated. Genes with assigned functions included those involved in phage morphogenesis, such as major capsid protein, head decoration protein, portal protein, prohead protease, baseplate hub protein, tail tube protein, tail terminator protein, and tape measure protein, and genes involved in DNA replication, such as DNA primase, NAD-dependent DNA ligase subunits A and B, DNA polymerase, and DNA helicase. Lysis genes coding for class III holin, endolysin with a peptidase domain, and an overlapping spanin pair were also found. Seabear shares 94.8% and 59.2% overall DNA sequence identity with Salmonella phage Stitch (GenBank accession number KM236244) (20) and phage T5 (GenBank accession number AY543070), respectively, as determined by BLASTn.

**Data availability.** The genome sequence of phage Seabear was submitted to GenBank as accession number MK728824. The associated BioProject, SRA, and BioSample accession numbers are PRJNA222858, SRR8771452, and SAMN11234143, respectively.

**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 Initiative Supported by Texas A&M University and Texas AgriLife, and from the Department of Biochemistry and Biophysics at Texas A&M University.

We are grateful for the advice and support of the CPT staff and the Texas A&M University Microscopy and Imaging Center.

This announcement was prepared in partial fulfillment of the requirements for BICH446 Phage Genomics, an undergraduate course at Texas A&M University.

**REFERENCES**

1. Antunes P, Mourão J, Campos J, Peixe L. 2016. Salmonellosis: the role of poultry meat. Clin Microbiol Infect 22:110–121. https://doi.org/10.1016/j.cmi.2015.12.004.
2. Fábrega A, Vila J. 2013. Salmonella enterica serovar Typhimurium skills to succeed in the host: virulence and regulation. Clin Microbiol Rev 26:308–341. https://doi.org/10.1128/CMR.00066-12.
3. Anderson CJ, Kendall MM. 2017. Salmonella enterica serovar Typhimurium strategies for host adaptation. Front Microbiol 8. https://doi.org/10.3389/fmicb.2017.01983.
4. Carvalho C, Costa AR, Silva F, Oliveira A. 2017. Bacteriophages and their derivatives for the treatment and control of food-producing animal infections. Crit Rev Microbiol 43:583–601. https://doi.org/10.1080/1040841X.2016.1271309.
5. Adams MK. 1959. Bacteriophages. Interscience Publishers, Inc., New York, NY.
6. Gill JJ, Berry JD, Russell WK, Lessor L, Escobar-Garcia DA, Hernandez MV, Kane A, Keene J, Maddox M, Martin R, Mohan S, Thorn AM, Russell DH, Young R. 2012. The Caulobacter crescentus phage phiCkC: genomics of a canonical phage. MBio Genomics 13:542. https://doi.org/10.1186/1471-2164-13-542.
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, 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.
8. 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.
9. Noguchi H, Taniguchi T, Inoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. Nucleic Acids Res 36:387–396. https://doi.org/10.1093/nar/gdn027.
10. 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.
11. 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.
12. The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:26599. https://doi.org/10.1093/nar/gky092.
13. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesse S, Quill AF, Sangrador-Vegas A, Scherermetjew 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.
14. Junker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci 12:1652–1662. https://doi.org/10.1111/j.1600-0839.2003.030703.
15. 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.
16. 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.
17. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes H, Elzik CG, Lewis SE. 2013. Web Apollo: a Web-based genomic annotation editing platform. Genome Biol 14:893. https://doi.org/10.1186/gb-2013-14-8-893.
18. Garneau JR, Depardieu F, Fortier LC, 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 7:28292. https://doi.org/10.1038/s41598-017-07910-5.
19. McClelland M, Sandersson KE, Spieth J, Clifton SW, Lareille P, Courtney L, Holmes H, Elzik CG, Lewis SE. 2013. Web Apollo: a Web-based genomic annotation editing platform. Genome Biol 14:893. https://doi.org/10.1186/gb-2013-14-8-893.
20. Grover JM, Luna AJ, Wood TL, Chamakura KR, Kuty Everett GF. 2015. Complete genome of Salmonella enterica serovar Typhimurium TS-like siphophage Stitch. Genome Announc 3:e01143-14. https://doi.org/10.1128/genomeA.01143-14.