



Complete Genome Sequence of *Enterococcus faecalis* Siphophage Sigurd

Adam Tomaszewski,^a Daniel Mora,^b James Clark,^{a,b} Tram Le,^{a,b}  Ry Young,^{a,b}  Mei Liu^{a,b}

^aDepartment of Biochemistry and Biophysics, Texas A&M University, College Station, Texas, USA

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

ABSTRACT *Enterococcus faecalis* is associated with antibiotic-resistant infections, and this study presents *E. faecalis* siphophage Sigurd. The 41,811-bp Sigurd genome is divided into two arms defined by long convergent predicted transcription units that are separated by a bidirectional rho-independent terminator. Sigurd has a small terminase that is closely related to *Bacillus subtilis* *cos* phage phi105.

Enterococcus faecalis is a common component of the human gut microbiome that is known to cause antibiotic-resistant nosocomial infections (1). The ineffectiveness of antibiotic treatments for infections caused by this pathogen necessitates the development of alternate strategies such as phage therapy. This study presents the annotated genome of phage Sigurd, which can infect *E. faecalis*.

Phage Sigurd was isolated in September 2019 from a wastewater influent sample collected in Port Aransas, Texas. It was propagated with soft agar overlay methods (2) using an *Enterococcus faecalis* isolate from mouse feces (strain Sor, provided by the University of California, San Diego [UCSD]) grown on brain heart infusion medium (Difco) at 37°C with aeration, and it showed clear plaque morphology. Phage DNA was purified using the Promega Wizard DNA cleanup system as described (3). DNA sequencing libraries were prepared as 300-bp inserts using a Swift 2S Turbo kit and were sequenced on an Illumina MiSeq system with paired-end 150-bp reads using 300-cycle v2 chemistry. The 235,068 individual raw reads were quality controlled using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed with FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit), yielding 164,226 reads. The reads were assembled using SPAdes v3.5.0 (4) to produce a contig with 66-fold coverage. The contig end sequences were completed by PCR (with primers 5'-TAGGCAACTGATGGCAAAC-3' and 5'-AGCGTTTTTCAGTCGCCAAT-3') using phage genomic DNA as the template, with Sanger sequencing of the resulting PCR product. The Center for Phage Technology (CPT) Galaxy-Apollo phage annotation platform (<https://cpt.tamu.edu/galaxy-pub>) was used to annotate the genome (5–7), utilizing GLIMMER v3 (8) and MetaGeneAnnotator v1.0 (9) for gene prediction, ARAGORN v2.36 (10) and tRNAScan-SE v2.0 (11) for tRNA prediction, and TransTermHP (12) for identification of rho-independent terminators. The prediction of gene functions utilized InterProScan v5.48 (13), BLAST v2.9.0 (14), TMHMM v2.0 (15), HHPred (16), Lipop v1.0 (17), and SignalP v5.0 (18). BLAST searching was performed against the NCBI nonredundant and Swiss-Prot databases (19). Genome-wide nucleotide similarity to top BLAST hits was calculated by progressiveMauve v2.4 (20). PhageTerm was used to predict phage termini (21). All tools were run with default settings.

Phage Sigurd has a genome of 41,811 bp, with a GC content of 34.5% and a coding density of 91.5%. It has 76 annotated genes, of which 27 have predicted functions. Phage Sigurd likely uses *cos*-type packaging, based on the fact that its small terminase subunit is closely related to that of the well-studied *Bacillus subtilis* *cos* phage phi105 (22). The exact location of the *cos* site, however, could not be predicted by PhageTerm. The genome of Sigurd can be divided into two arms defined by long convergent predicted transcription

Editor Kenneth M. Stedman, Portland State University

Copyright © 2022 Tomaszewski 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 Mei Liu, meiliu@tamu.edu.

The authors declare no conflict of interest.

Received 9 February 2022

Accepted 10 March 2022

Published 28 March 2022

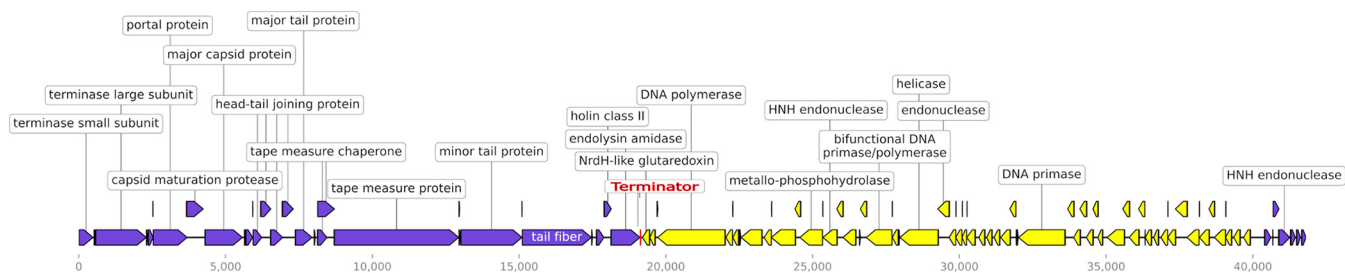


FIG 1 Genome map of phage Sigurd. Genes in purple are transcribed on the plus strand, and genes in yellow are transcribed on the minus strand. The predicted terminator location is indicated in red.

units that are separated by a large bidirectional rho-independent terminator. Convergent transcription is presumed from the gene orientation at this location. All of the genes on either side of this point are on the same coding strand and are thus presumed to be cotranscribed, as shown in the genome map (Fig. 1). The left arm has many predicted virion structural and lysis genes, all on the plus strand. The right arm of the genome, encoded mostly on the minus strand, is dominated by novel hypothetical genes. A predicted bifunctional DNA primase/polymerase (gp35) and two HNH endonucleases were identified, one of which is near the end of the genome and may play a role in *cos* cleavage (23). As determined by progressiveMauve, Sigurd shares >70% nucleotide similarity with other *Enterococcus* phages such as vB_EfaS_Ef6.1 (GenBank accession number [MK721187](https://doi.org/10.1371/journal.pone.0000582)) and vB_EfaS_Ef6.4 (GenBank accession number [MK721190](https://doi.org/10.1371/journal.pone.0000582)). Sigurd was determined to have a siphophage morphology via transmission electron microscopy, but the image is not shown due to poor image quality.

Data availability. Sigurd was deposited in GenBank with accession number [MZ326865](https://doi.org/10.1371/journal.pone.0000582). The associated BioProject, SRA, and BioSample accession numbers are [PRJNA222858](https://doi.org/10.1371/journal.pone.0000582), [SRR14095245](https://doi.org/10.1371/journal.pone.0000582), and [SAMN18509620](https://doi.org/10.1371/journal.pone.0000582), respectively.

ACKNOWLEDGMENTS

Funding was provided by the National Science Foundation (awards EF-0949351 and DBI-1565146) and the CPT. Yi Duan at UCSD provided the *E. faecalis* strain used in this study.

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

REFERENCES

- McBride SM, Fischetti VA, LeBlanc DJ, Moellering RC, Jr, Gilmore MS. 2007. Genetic diversity among *Enterococcus faecalis*. *PLoS One* 2:e582. <https://doi.org/10.1371/journal.pone.0000582>.
- Wommack KE, Williamson KE, Helton RR, Bench SR, Winget DM. 2009. Methods for the isolation of viruses from environmental samples. *Methods Mol Biol* 501:3–14. https://doi.org/10.1007/978-1-60327-164-6_1.
- 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.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>.
- Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning 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>.
- Dunn NA, Unni DR, Diesch C, Munoz-Torres M, Harris NL, Yao E, Rasche H, Holmes IH, Elsik CG, Lewis SE. 2019. Apollo: democratizing genome annotation. *PLoS Comput Biol* 15:e1006790. <https://doi.org/10.1371/journal.pcbi.1006790>.
- Ramsey J, Rasche H, Maughmer C, Criscione A, Mijalis E, Liu M, Hu JC, Young R, Gill JJ. 2020. Galaxy and Apollo as a biologist-friendly interface for high-quality cooperative phage genome annotation. *PLoS Comput Biol* 16:e1008214. <https://doi.org/10.1371/journal.pcbi.1008214>.
- 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>.
- 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>.
- 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>.
- Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol* 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.
- 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 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>.
14. 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>.
 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. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI Bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 430:2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007>.
 17. Juncker 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/10.1110/ps.0303703>.
 18. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>.
 19. UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 46:2699. <https://doi.org/10.1093/nar/gky092>.
 20. 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>.
 21. Garneau JA-O, Depardieu F, Fortier LC, Bikard DA-O, Monot MA-O. 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>.
 22. Scher BM, Dean DH, Garro AJ. 1977. Fragmentation of *Bacillus* bacteriophage ϕ 105 DNA by complementary single-stranded DNA in the cohesive ends of the molecule. *J Virol* 23:377–383. <https://doi.org/10.1128/JVI.23.2.377-383.1977>.
 23. Kala S, Cumby N, Sadowski PD, Hyder BZ, Kanelis V, Davidson AR, Maxwell KL. 2014. HNH proteins are a widespread component of phage DNA packaging machines. *Proc Natl Acad Sci U S A* 111:6022–6027. <https://doi.org/10.1073/pnas.1320952111>.