AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Escherichia coli* Phage Paul

Ashley Holt,[a] Robert Saldana,[a] Russell Moreland,[a] Jason J. Gill,[a] Mei Liu,[a] Jolene Ramsey[a]

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

ABSTRACT *Escherichia coli* is both a commensal and a pathogen in humans and other animals. Here, we describe the isolation of *E. coli* strain 4s bacteriophage Paul. The complete 79,429-bp genome was annotated and demonstrates similarity with phieco32viruses, as does its prolate podophage morphology.

Escherichia coli is a commensal bacterial inhabitant of the intestines, with pathogenic groups that cause human disease (1). *E. coli* strain 4s is a commensal isolate collected from horse feces and has an O-antigen component of the lipopolysaccharide known to affect susceptibility to phage (2). Here, we present the complete, annotated genome sequence of the *E. coli* 4s prolate podophage Paul.

Bacteriophage Paul was isolated from a filtered (0.2-μm-pore-size) water sample collected at Wolf Pen Creek in College Station, TX. The phage was propagated on *E. coli* 4s aerobically at 37°C in Luria-Bertani broth (BD Difco) using the soft-agar overlay methods described by Adams (3). DNA was purified with the modified Promega Wizard DNA clean-up system shotgun library preparation protocol (4), prepared as Illumina TruSeq Nano low-throughput libraries, and sequenced on an Illumina MiSeq platform with paired-end 250-bp reads using V2 500-cycle chemistry. The 2,820,474 reads in the phage index were quality controlled using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequence reads were then trimmed using the FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). The genome was assembled into a single contig with 1,429.4-fold coverage using SPAdes v3.5.0, with default parameters, and was confirmed to be complete by Sanger sequencing of a PCR product amplified off the raw contig ends (forward primer, 5′-CGTCGGCAATATCGTCTACTTT-3′, and reverse primer, 5′-AACAGCCTTACAATCCCTTACTG-3′) (5). Structural annotations were performed with GLIMMER v3.0 and MetaGeneAnnotator v1.0, and tRNA sequences were detected with ARAGORN v2.36 (6–8). Rho-independent termination sites were annotated using TransTermHP v2.09 (9). Gene functions were predicted using InterProScan v5.33-72, BLAST v2.2.31, and TMHMM v2.0, with default settings (10–12). BLAST searches were executed against the NCBI nonredundant and UniProtKB Swiss-Prot/TrEMBL databases with a 0.001 maximum expectation value (13). Structural predictions were done with the HHSuite v3.0 tool HHpred (multiple-sequence alignment [MSA] generation with HHblits using the ummiclus30_2018_08 database and modeling with the PDB_mmCIF70 database) (14). Genome-wide DNA sequence similarity was calculated by progressiveMauve v2.4.0, with default parameters (15). The annotation tools were accessed in the Galaxy and Web Apollo tools hosted by the Center for Phage Technology (https://cpt.tamu.edu/galaxy-pub) (16, 17) and run with default parameters (unless otherwise stated). The morphology of phage Paul was determined from samples negatively stained with 2% (wt/vol) uranyl acetate and viewed by transmission electron microscopy at the Texas A&M Microscopy and Imaging Center (18).

Paul is a 79,429-bp prolate podophage with 42.0% G+C content and 91.4% coding density. Structural annotations yielded 133 predicted protein-coding genes and a single tRNA gene. By BLASTp, Paul shares 113 proteins similar to those of enterobacteria phage phiEco32 (GenBank accession number EU330206), a 77-kb prolate podophage

Citation Holt A, Saldana R, Moreland R, Gill JJ, Liu M, Ramsey J. 2019. Complete genome sequence of *Escherichia coli* phage Paul. Microbiol Resour Announc 8:e01093-19. https://doi.org/10.1128/MRA.01093-19.

Editor John J. Dennehy, Queens College

Address correspondence to Jolene Ramsey, jolenerr@tamu.edu.

Received 3 September 2019
Accepted 18 September 2019
Published 10 October 2019

isolated against *E. coli* from cattle with acute mastitis (19). At the nucleotide level, Paul is most similar to other *Phieco32virus* members, including phage vB_EcoP_SU10 (82.24%, KM044272), phiEco32 (82.03%, EU330206), enterobacteria phage NJ01 (81.67%, JX867715), and *Escherichia* phage 172-1 (80.63%, KP308307). PhageTerm predicted 193-bp direct terminal repeats, and the assembled genome was reopened at the left terminal repeat boundary, syntenic with phiEco32 (20).

Data availability. The genome sequence and associated data for phage Paul were deposited under GenBank accession number MN045231, BioProject number PRJNA222858, SRA number SRR8892204, and BioSample number SAMN11411459.

ACKNOWLEDGMENTS

This work was supported by funding from the National Science Foundation (award DBI-1565146). Additional support came from the Center for Phage Technology (CPT), an Initial University Multidisciplinary Research Initiative supported by Texas A&M University and Texas AgriLife, and from the Department of Biochemistry and Biophysics at Texas A&M University.

We thank A. Letarov for the kind gift of the *Escherichia coli* strain 4s. 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

1. Fratamico PM, DebRoy C, Liu Y, Needleman DS, Baranzoni GM, Feng P. 2016. Advances in molecular serotyping and subtyping of Escherichia coli. Front Microbiol 7:644. https://doi.org/10.3389/fmicb.2016.00644.
2. Knirel YA, Prokhorov NS, Shashkov AS, Ovchinnikova OG, Zdorovenko EL, Liu B, Kostryukova ES, Larin AK, Golomidova AK, Letarov AV. 2015. Variations in O-antigen biosynthesis and O-acetylation associated with altered phage sensitivity in Escherichia coli 4s. J Bacteriol 197:905–912. https://doi.org/10.1128/JB.02398-14.
3. Adams MH. 1956. Bacteriophages. Interscience Publishers, Inc., New York, NY.
4. 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.
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. https://doi.org/10.1089/cmb.2012.0021.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
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. Krogh A, Larsson B, Heijne von 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.
13. The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049.
14. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding 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.
15. 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.
16. 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, Hiltemann 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.
17. 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.
18. 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.
19. Savalia D, Westblade LF, Goel M, Florens L, Kemp P, Akulenko N, Pavlova O, Padovan JC, Chait BT, Washburn MP, Ackermann H-W, Mushegian A, Gabisonia T, Molineux I, Severinov K. 2008. Genomic and proteomic analysis of phiEco32, a novel Escherichia coli bacteriophage. J Mol Biol 377:774–789. https://doi.org/10.1016/j.jmb.2007.12.077.
20. 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.