



Draft Genome Sequence of *Escherichia coli* Phage CMSTMSU, Isolated from Shrimp Farm Effluent Water

Lelin Chinnadurai,^a Thirumalaikumar Eswaramoorthy,^a Abinaya Paramachandran,^a Sayan Paul,^b Rashmi Rathy,^b Arun Arumugaperumal,^b Sudhakar Sivasubramaniam,^b Citarasu Thavasimuthu^a

^aCentre for Marine Science and Technology, Manonmaniam Sundaranar University at Rajakkamangalam, Kanyakumari, Tamilnadu, India

^bDepartment of Biotechnology, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India

ABSTRACT The *Escherichia coli* phage CMSTMSU was isolated from shrimp farm effluent water in Ramanathapuram, India. The phage exhibited lytic activity against both *E. coli* and the fish pathogen *Pseudomonas aeruginosa*. Here we report the draft genome sequence, assembly, and annotation of the isolated CMSTMSU phage. This genome resource can be used to utilize the phage as a crucial biocontrol agent in the fish aquaculture sector.

Bacteriophages are viruses that infect bacteria. They are abundant in natural systems and are considered crucial factors in controlling bacterial populations (1). Phages also have the potential to regulate bacterial diseases of fish in aquatic environments by removing the fish pathogens (2). This study reports the genome sequence, assembly, and annotation of the *Escherichia coli* phage CMSTMSU. The phage was isolated from a wastewater sample obtained from a shrimp farm located in Ramanathapuram, India. It was detected with the soft agar overlay method using log-phase *E. coli* cells as the host. The isolated CMSTMSU phage also exhibited lytic activity against the fish pathogen *Pseudomonas aeruginosa*.

The *E. coli* phage CMSTMSU was purified following the protocol reported by Mullan (see <https://www.dairyscience.info/index.php/isolation-and-purification-of-bacteriophages.html>). Then, the genomic DNA was extracted with the phenol-chloroform extraction method (3). The DNA library was prepared with the NEBNext Ultra II DNA library prep kit (New England Biolabs, USA). The whole-genome sequencing was performed with MinION Mk1b (Oxford Nanopore Technologies, UK) using the SpotON flow cell (FLO-MIN106) (4), and base calling was performed with Albacore version 2.1.3 at Genotypic Technology Pvt Ltd (Bangalore, India). We obtained 88,676 reads from the bar-coded library with the Nanopore sequencer with an average read length of 3.4 kb and an N_{50} length of 6,531 bp. The quality of the reads was analyzed with FastQC software version 0.11.5 (5). The base-called raw reads were used for *de novo* assembly with the Canu algorithm (6). The Canu assembly generated a single contig of 386.4 kb, which has a GC content of 35.6%. The contig underwent a BLAST search against the NCBI virus nonredundant (nr) database with the BLASTN algorithm with an E value threshold of $1E-5$, and we found that it has an 83% sequence similarity with *Escherichia* phages PBECO 4, vB_Eco_slurp01, and 121Q.

The draft genome of *E. coli* phage CMSTMSU was annotated with the RAST annotation server version 2.0 (<http://rast.nmpdr.org>) (7), GeneMarkS version 4.28 (<http://exon.gatech.edu/GeneMark/genemarks.cgi>) (8), and GLIMMER version 3.02 (<https://ccb.jhu.edu/software/glimmer/>) (9) gene prediction tools. The data obtained from the RAST annotation identified 767 protein-coding genes, and among them, 715 (91%) genes were identified from a BLAST search against the NCBI virus database with the BLASTP algorithm. The gene ontology (GO) and KEGG pathway annotations of the protein-

Received 24 July 2018 **Accepted** 10 September 2018 **Published** 11 October 2018

Citation Chinnadurai L, Eswaramoorthy T, Paramachandran A, Paul S, Rathy R, Arumugaperumal A, Sivasubramaniam S, Thavasimuthu C. 2018. Draft genome sequence of *Escherichia coli* phage CMSTMSU, isolated from shrimp farm effluent water. *Microbiol Resour Announc* 7:e01034-18. <https://doi.org/10.1128/MRA.01034-18>.

Editor Irene L. G. Newton, Indiana University Bloomington

Copyright © 2018 Chinnadurai 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 Citarasu Thavasimuthu, citarasu@msuniv.ac.in.

coding genes were performed with the Blast2GO (<https://www.blast2go.com/>) functional annotation software (10). Of the 715 BLAST-annotated genes, 190 genes were assigned to 423 GO terms with ATP binding (45 genes) and nucleic acid phosphodiester bond hydrolysis (32 genes), and these were the most highly represented GO terms in the data set. We mapped 117 genes with 12 KEGG metabolic pathways, among which the pathways associated with purine metabolism (37 genes) and pyrimidine metabolism (26 genes) were the most dominant in the genome data set. Simultaneously, the annotations with the GeneMarkS and GLIMMER gene prediction tools predicted 891 and 938 protein-coding genes, respectively. Among these predicted genes, 599 genes were common to all three databases, whereas 115, 12, and 115 genes showed an overlap between RAST and GLIMMER, RAST and GeneMarkS, and GeneMarkS and GLIMMER, respectively. In addition, we identified 6 tRNA genes with a GC content range from 48.6% to 58.4% with the ARAGORN version 1.2.38 program (11). This genome draft sequence can be used as a potential resource to utilize the phage species as a biocontrol agent of antibiotics against fish pathogens.

Data availability. The raw sequence reads have been submitted to the NCBI SRA under the accession number [SRP158495](https://www.ncbi.nlm.nih.gov/sra/SRP158495), and the draft genome sequence of *Escherichia coli* phage CMSTMSU has been deposited in NCBI GenBank under the accession number [MH494197](https://www.ncbi.nlm.nih.gov/genbank/MH494197).

ACKNOWLEDGMENTS

This work was financially supported by the Basic Science Research (BSR), University Grants Commission (UGC), New Delhi, Government of India (grant F.25-1/2013-2014 [BSR]/7-374/2012 [BSR] dated 30 May 2014). A.A. is supported by the DBT (grant DBT/2015/SJRI/447).

REFERENCES

1. Abedon ST. 2008. Bacteriophage ecology: population growth, evolution, and impact of bacterial viruses, vol 15. Cambridge University Press, Cambridge, United Kingdom.
2. Pereira C, Silva YJ, Santos AL, Cunha Â, Gomes NCM, Almeida A. 2011. Bacteriophages with potential for inactivation of fish pathogenic bacteria: survival, host specificity and effect on bacterial community structure. *Mar Drugs* 9:2236–2255. <https://doi.org/10.3390/md9112236>.
3. Elmaghraby I, Carimi F, Sharaf A, Marei EM, Hammad AMM. 2015. Isolation and identification of *Bacillus megaterium* bacteriophages via AFLP technique. *Curr Res Bacteriol* 8:77. <https://doi.org/10.3923/crb.2015.77.89>.
4. Loose M, Malla S, Stout M. 2016. Real-time selective sequencing using nanopore technology. *Nat Methods* 13:751. <https://doi.org/10.1038/nmeth.3930>.
5. Babraham Bioinformatics. 2011. FastQC: a quality control tool for high throughput sequence data. Babraham Institute, Cambridge, United Kingdom. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
6. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
7. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
8. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29:2607–2618. <https://doi.org/10.1093/nar/29.12.2607>.
9. 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>.
10. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>.
11. 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>.