UNIVERSIDAD SIMÓN BOLÍVAR | INSTITUTO DE POSGRADOS

Genomic Characterization of the Lytic Phage vB_EcoP_EcoN5 that infects Shiga toxin-producing Escherichia coli O157: H7

Cristian Alfredo Solano Castañeda
"**Magister en Genética"**
Antonio H. Acosta Hoyos
Dayan Lozano Solano

RESUMEN

Los bacteriófagos infectan y matan bacterias de forma selectiva y eficiente.La diseminación de cepas de *Escherichia coli* resistentes a los antibióticos volvió a atraer a los fagos como agentes terapéuticos potenciales. El objetivo de este estudio fue caracterizar un nuevo fago lítico, vB_EcoP_EcoN5, que infecta *Escherichia coli O157: H7* productora de toxina shiga donde se tomó muestras de aguas residuales en Barranquilla – Colombia, para asilar el bacteriófago mediante un método de enriquecimiento, el cual permitió determinar la susceptibilidad H7 y se secuenció el genoma completo; una vez obtenida la biblioteca final que fue secuenciada con el instrumento de celda de flujo MiSeq v3 (Illumina Inc.), para generar al menos un millón de lecturas se ensamblaron en un único contig con una cobertura de 14217 veces usando SPAdes, después que Trimmomatic eliminó las secuencias del adaptador se realizó la respectiva anotación utilizando RAST (Anotación Rápida Usando Tecnología de Subsistema) (http://rast.theseed.org/FIG/rast.cgi), que es un servicio totalmente automatizado para anotar genomas completos de arca y fagos, y asignar funciones a los genes (Inicialmente se ejecutó un BLAST (https://blast.ncbi.nlm.nih.gov), para encontrar genomas con alta identidad como referencia y se usó también vB_EcoP_Eco32 como genoma de referencia (ID de taxonomía: 490103, Genbank NC: 010324), en la plataforma RAST, el genoma de referencia con su taxonomía ID se utilizó para llevar a cabo el proceso de anotación y se obtuvo un archivo gbk; el cual fue visualizado por Uniprot UGENE mediante BLASTP se verificó manualmente la función de cada ORF y la base de datos Pfam. Para analizar la relación filogenética de vB_EcoP_EcoN5 se utilizaron secuencias de proteínas tanto de la cabeza como de la ADN polimerasa para la construcción de los árboles filogenéticos, así como las secuencias completas del genoma. Los árboles se construyeron con representantes del género *Kuravirus* y otros miembros de la familia *Podoviridae* según el último informe de la ICTV (Comité Internacional de Taxonomía de Virus, 2018), (https://talk.ictvonline.org/taxonomy), usando el método de unión de vecinos con un valor de arranque = 500 empleando MEGA versión X. El genoma tenía 76.083 pb de dsDNA de largo con un G + C de 42.09%. La anotación de EcoN5 reveló 128 ORF donde 40 ORF tenían una función predicha por el análisis de Blastp y HHPred, y 88 ORF no revelaron homología funcional aparente con otras proteínas, aquellos ORF que presentaron

alguna función específica se agruparon en 5 módulos, de los cuales 23 ORF pertenecen al módulo de regulación y replicación del ADN, 10 ORF, estructura de fagos y módulo de ensamblaje, 3 ORF, Módulo de lisis, 3 ORF, módulo de regulación del ciclo de vida y 1 proteína accesoria. EcoN5 codifica el genoma de un raro ARGN tRNA-anticodon UCU, lo que podría dar a este fago una ventaja de replicación durante el ciclo de infección. El análisis filogenético y la identidad de nucleótidos promedio colocaron a EcoN5 como un nuevo miembro del género *Kuravirus* dentro de la subfamilia *Sepvirinae*, familia *Podoviridae*, presentando una mayor similitud con el fago Eco32 (número de acceso de Genbank EU330206.1), 92.57%, y el fago 172-1 (número de acceso de Genbank KP308307.1), 92.46%, también EcoN5 comparte más del 90% de identidad con otros 5 miembros del género *Kuravirus* como son vB_EcoliP_Eco32, 172-1, 2311, Paul y SU10 según los datos obtenidos de los nucleótidos promedios (ANI). Este estudio avanza el conocimiento de los fagos de E. coli y proporciona un posible candidato para ser considerado para la terapia con fagos. La secuencia del genoma anotada del bacteriófago vB_EcoP_EcoN5 se depositó en la base de datos GenBank y está disponible con el número de acceso MN715356.

Palabras clave: Terapia con fagos, diversidad de fagos, resistencia a antibióticos, fagos de *E. coli*

ABSTRACT

Bacteriophages infect and kill bacteria selectively and efficiently. The spread of antibiotic-resistant Escherichia coli strains once again attracted phages as potential therapeutic agents.The objective of this study was to characterize a new lithic phage, vB_EcoP_EcoN5, that infects *Escherichia coli* O157: H7 producing shiga toxin where sewage samples were taken in Barranquilla - Colombia, to isolate the bacteriophage by an enrichment method, which allowed determine H7 susceptibility and the complete genome was sequenced; Once the final library that was sequenced with the MiSeq v3 flow cell instrument (illuminates Inc.) was obtained, to generate at least one million readings they were assembled in a single contig with a coverage of 14217 times using spades after Trimmomatic removed the adapter sequences, the respective annotation was performed using RAST (Rapid Annotation Using Subsystem Technology) (http://rast.theseed.org/FIG/rast.cgi), which is a fully automated service to annotate complete ark genomes and phages, and assign functions to the genes (Initially a BLAST) (https://blast.ncbi.nlm.nih.gov) was run, to find genomes with high identity as a reference and vB_EcoP_Eco32 was also used as a reference genome (ID of taxonomy: 490103, Genbank NC: 010324) on the RAST platform, the reference genome with its ID taxonomy was used to carry out the annotation process and a gbk file was obtained; which was visualized by Uniprot UGENE via BLASTP, the function of each ORF and the Pfam database were verified manually. To analyze the phylogenetic relationship of vB_EcoP_EcoN5 protein sequences from both the head and DNA polymerase were used for the construction of phylogenetic trees, as well as the complete genome sequences. The trees were built with representatives of the genus *Kuravirus* and other members of the *Podoviridae* family according to

the latest report of the ICTV (International Committee of Taxonomy of Virus, 2018), (https://talk.ictvonline.org/taxonomy), using the neighbor binding method with a starting value = 500 using MEGA version X. The genome was 76,083 bp of dsDNA long with a G + C of 42.09%. The EcoN5 annotation revealed 128 ORF where 40 ORF had a function predicted by the Blastp and HHPred analysis, and 88 ORF did not reveal apparent functional homology with other proteins, those ORF that presented some specific function were grouped into 5 modules, of which 23 ORF belong to the DNA regulation and replication module, 10 ORF, phage structure and assembly module, 3 ORF, lysis module, 3 ORF, life cycle regulation module and an accessory protein. EcoN5 encodes the genome of a rare tRNA-anticodon ARGN UCU, which could give this phage an advantage of replication during the infection cycle. The phylogenetic analysis and the average nucleotide identity placed EcoN5 as a new member of the Kuravirus genus within the *Sepvirinae* subfamily, family *Podoviridae*, presenting a greater similarity with the phage Eco32 (Genbank accession number EU330206.1), 92.57%, and phage 172-1 (Genbank accession number KP308307.1), 92.46%, also EcoN5 shares more than 90% identity with 5 other members of the Kuravirus genus such as vB_EcoliP_Eco32, 172-1, 2311, Paul and SU10 according the data obtained from the average nucleotides (ANI). This study advances knowledge of E. coliphages and provides a possible candidate to be considered for phage therapy. The annotated genome sequence of bacteriophage vB_EcoP_EcoN5 was deposited in the GenBank database and is available under accession number MN715356.

KeyWords: Phage therapy, Phage diversity, antibiotic-resistance, *E. coli* phages

REFERENCIAS

1. Klumpp J, Fouts DE, Sozhamannan S. Next generation sequencing technologies and the changing landscape of phage genomics. Bacteriophage [Internet]. 2012 Jul 1;2(3):190–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23275870

2. Philipson CW, Voegtly LJ, Lueder MR, Long KA, Rice GK, Frey KG, et al. Characterizing Phage Genomes for Therapeutic Applications. Viruses [Internet]. 2018 Apr 10;10(4):188. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29642590

3. Chen Y, Batra H, Dong J, Chen C, Rao VB, Tao P. Genetic Engineering of Bacteriophages Against Infectious Diseases. Front Microbiol [Internet]. 2019;10:954. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2019.00954

4. Clokie MR, Millard AD, Letarov A V, Heaphy S. Phages in nature. Bacteriophage [Internet]. 2011 Jan;1(1):31–45. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21687533

5. Batinovic S, Wassef F, Knowler SA, Rice DTF, Stanton CR, Rose J, et al. Bacteriophages in natural and artificial environments. Pathogens. 2019;8(3):1–19.

6. Haq IU, Chaudhry WN, Akhtar MN, Andleeb S, Qadri I. Bacteriophages and their implications on future biotechnology: a review. Virol J [Internet]. 2012 Jan 10;9:9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22234269

7. Campbell A. The future of bacteriophage biology. Nat Rev Genet [Internet]. 2003;4(6):471–7. Available from: https://doi.org/10.1038/nrg1089

8. Jurczak-Kurek A, Gąsior T, Nejman-Faleńczyk B, Bloch S, Dydecka A, Topka G, et al. Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage. Sci Rep [Internet]. 2016;6(1):34338. Available from: https://doi.org/10.1038/srep34338

9. Torres AG, Amaral MM, Bentancor L, Galli L, Goldstein J, Kruger A, et al. Recent Advances in Shiga Toxin-Producing Escherichia coli Research in Latin America. Microorganisms. 2018 Sep;6(4).

10. Kropinski AM, Mazzocco A, Waddell TE, Lingohr E, Johnson RP. Enumeration of bacteriophages by double agar overlay plaque assay. Methods Mol Biol [Internet]. 2009 Jan [cited 2015 Mar 9];501:69–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19066811

11. Kutter E. Phage host range and efficiency of plating. Methods Mol Biol. 2009;501:141–9.

12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May;19(5):455–77.

13. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug;30(15):2114–20.

14. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics [Internet]. 2008;9(1):75. Available from: https://doi.org/10.1186/1471-2164-9-75

15. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep;25(17):3389–402.

16. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012 Apr;28(8):1166–7.

17. Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genet. 1993 Mar;3(3):266–72.

18. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019 Jan;47(D1):D427–32.

19. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005 Apr;21(7):951–60.

20. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kubler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 2018 Jul;430(15):2237–43.

21. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 2011 Jan;39(Database issue):D225-9.

22. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017 Jan;45(D1):D200–3.

23. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016 Jul;33(7):18704.

24. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics [Internet]. 2011/01/28. 2011 Apr 1;27(7):1009–10. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21278367

25. Aguirre-López B, Cabrera N, de Gómez-Puyou MT, Perez-Montfort R, Gómez-Puyou A. The importance of arginine codons AGA and AGG for the expression in E. coli of triosephosphate isomerase from seven different species. Biotechnol reports (Amsterdam, Netherlands) [Internet]. 2017 Jan 8;13:42–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28352562

26. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016 Jul;44(W1):W54-7.

27. Figueras MJ, Beaz-Hidalgo R, Hossain MJ, Liles MR. Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. Genome Announc [Internet]. 2014 Dec 4;2(6):e00927-14. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25477398

28. Mirzaei MK, Eriksson H, Kasuga K, Haggård-Ljungquist E, Nilsson AS. Genomic, Proteomic, Morphological, and Phylogenetic Analyses of vB_EcoP_SU10, a Podoviridae Phage with C3 Morphology. PLoS One [Internet]. 2014 Dec 31;9(12):e116294. Available from: https://doi.org/10.1371/journal.pone.0116294

29. Fawaz M V, Topper ME, Firestine SM. The ATP-grasp enzymes. Bioorg Chem. 2011 Dec;39(5–6):185–91.

30. Iyer LM, Abhiman S, Maxwell Burroughs A, Aravind L. Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: synthesis of novel metabolites and peptide modifications of proteins. Mol Biosyst. 2009 Dec;5(12):1636—1660.

31. Savalia D, Westblade LF, Goel M, Florens L, Kemp P, Akulenko N, et al. Genomic and proteomic analysis of phiEco32, a novel Escherichia coli bacteriophage. J Mol Biol [Internet]. 2008/01/11. 2008 Mar 28;377(3):774–89. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18294652