Manuscript No: VIRUS-D-21-00729R1

Genome analysis of a new *Escherichia* phage vB_EcoM_C2-3 with lytic activity against multidrug-resistant *Escherichia coli*

Ana C. Maganha de Almeida Kumlien[1,2], Clara Pérez-Vega[1,2], Edgar González-Villalobos [1,2], Carles M. Borrego[1,3], José Luis Balcázar[1,2,*]

[1] Catalan Institute for Water Research (ICRA), Girona, Spain

[2] University of Girona, Girona, Spain

[3] Group of Molecular Microbial Ecology, Institute of Aquatic Ecology, University of Girona, Girona, Spain

*Corresponding author: *E-mail address:* jlbalcazar@icra.cat (J.L. Balcázar)

Abstract

In this study, we present the complete, annotated genome of a new member of the *Tequatrovirus* (T4-like) genus, *Escherichia* phage vB_EcoM_C2-3. This phage has an isometric head (92 nm in diameter) and a contractile tail (114 nm in length). Its genome consists of a linear, double-stranded DNA of 167,069 bp with an average G+C content of 35.3%. There are 267 predicted genes, of which 125 encode functional proteins, including those for DNA replication, transcription and packaging, phage morphogenesis and cell lysis. Neither genes involved in the regulation of lysogeny nor antibiotic resistance genes were identified. Based on our results, its genomic features provide valuable insights into the use of a potential biocontrol agent, as *Escherichia* phage vB_EcoM_C2-3 exhibited lytic activity against *E. coli*, including multidrug-resistant strains.

Keywords: Bacteriophage; T4-like phages; *Escherichia coli*; Genomic analysis

Escherichia coli is a Gram-negative bacterium, which is a member of the normal microbiota of humans and animals. Although most *E. coli* strains are harmless, some strains are implicated in several diseases, including diarrhea, septicemia, and urinary tract infections (Bai et al., 2020; van Hoffen et al., 2021). These infections are usually treated with antibiotics but most strains are becoming increasingly resistant to them, thereby limiting the effectiveness of antibiotic therapy (Bunduki et al., 2021).

Bacteriophages have shown their potential as alternative biocontrol agents to antibiotics (Chegini et al., 2021; González-Villalobos et al., 2021). Bacteriophages, or simply phages, are viruses that infect and frequently kill bacteria. In the framework of a previous study conducted on a local wastewater treatment plant, we isolated a phage exhibiting lytic activity against multi-drug resistant *E. coli* strains previously isolated from raw sewage samples, as well as against other reference strains of *E. coli* (e.g., ATCC 700078, BAA-2730, and HV1735). The aim of this study was therefore to describe its complete genome sequence and its taxonomic position.

Escherichia phage vB_EcoM_C2-3 was isolated from a sewage sample by enriching with *E. coli* strain C2 previously adjusted to an OD_{600} of 0.2 (1.6×10^8 CFU ml^{-1}) in tryptic soy broth and incubating at 37 °C overnight. The phage was then propagated three times to ensure purity following the process previously described (Kropinsky et al., 2009; González-Villalobos et al., 2021). *E. coli* strain C2 was chosen as a host because it exhibited resistance to several antimicrobial classes, including cephalosporins (cefotaxime), fluoroquinolones (ciprofloxacin), polymyxins (colistin), and sulfonamides (sulfamethoxazole), according to the criteria established by the Clinical and Laboratory Standards Institute (CLSI, 2016). Phage morphology was observed by transmission electron microscopy according to the previously described procedure (Wang et al., 2018), which revealed that *Escherichia* phage vB_EcoM_C2-3 has an isometric head of 92 nm

in diameter and possesses a contractile tail of 114 nm in length (Fig. 1). These features suggest that it belongs to the *Myoviridae* family. Subsequently, the phage lysate was treated with DNase I (Thermo Fisher Scientific, Waltham, MA) to remove cellular DNA and phage DNA was further purified by proteinase K treatment to digest capsid proteins, followed by a standard phenol-chloroform extraction and ethanol precipitation (Sambrook and Russell, 2006). After the purification process, the concentration of phage DNA was determined using the Qubit fluorometer (Thermo Fisher Scientific), and its quality was evaluated using agarose gel electrophoresis. The sequencing library was then prepared by random DNA fragmentation using Illumina TruSeq DNA PCR-free kit, followed by 5' and 3' adapter ligation. Approximately 20 million 2 × 150-bp paired-end reads were generated using the Illumina NovaSeq 6000 platform, according to the manufacturer's instructions. The raw reads were subjected to FastQC for quality check, followed by sequence trimming using Trimmomatic with default settings (Bolger et al., 2014). The resultant high-quality reads were assembled using SPAdes v3.15.3 (Bankevich et al., 2012), and open reading frames (ORFs) were predicted using Prokka v1.14.5 (Seemann, 2014) and PHASTER (Arndt et al., 2016). The predicted ORFs were subjected to a homology search using the NCBI protein BLAST server (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The phage packaging mechanisms and genome termini were predicted using PhageTerm (Garneau et al., 2017). Moreover, the presence of tRNAs was determined using the tRNAscan-SE search server (Lowe and Chan, 2016), whereas antibiotic resistance genes were examined on the ResFinder 4.1 database (Zankari et al., 2012). A map of the phage genome was generated using CGview server (Petkau et al., 2010), and genomic comparison with closely related phages was performed using the Easyfig software (Sullivan et al., 2011).

The complete genome sequence of *Escherichia* phage vB_EcoM_C2-3 was deposited in the GenBank database under the accession number OK076929. The genome of this phage consists of a linear, double-stranded DNA of 167,069 bp with an average G+C content of 35.3%, which is much lower than that of its host *E. coli* (50.6%). The genome was scanned for ORFs of 100 bp or longer, and the search resulted in 267 predicted genes with an average length of 588 bp, which occupy 93.9% of its genome (Fig. 2). No unique proteins were detected, and it was possible to predict a function for 125 ORFs. The functionally predicted gene products were grouped into three modules: DNA replication and transcription, DNA packaging and phage morphogenesis, and cell lysis (Fig. 3). Moreover, we identified eleven tRNA genes, namely: tRNA-Arg, tRNA-Asn, tRNA-Gln, tRNA-Gly, tRNA-His, tRNA-Leu, tRNA-Met, tRNA-Pro, tRNA-Ser, tRNA-Thr, tRNA-Tyr. Despite these tRNA genes may both facilitate phage integration and contribute to higher virulence (Miller et al., 2003; Bailly-Bechet et al., 2007), the reasons why some phages contain tRNAs still remain unknown. Genome analysis also revealed that this phage is virulent and does not encode any genes associated with lysogeny, as well as any genes encoding antibiotic resistance. PhageTerm predicted that *Escherichia* phage vB_EcoM_C2-3 uses a headful packaging mechanism without either a preferred *pac* site or terminal cohesive ends.

Homology searches revealed that *Escherichia* phage vB_EcoM_C2-3 is a new member of the *Tequatrovirus* (T4-like) genus, with 97.7% sequence identity over 97.0% coverage of the *Escherichia* phage D5505 (GenBank accession number MK327929.1), under the *Tevenvirinae* subfamily within the *Myoviridae* family. *Escherichia* phage vB_EcoM_C2-3 also shared 97.2 and 94.5% sequence identity over 96.0 and 88.0% coverage of *Escherichia* phage vB_EcoM_JB75 (GenBank accession number MH355584.1) and *Escherichia* phage T4 (GenBank accession number AF158101.6), respectively. Although

most predicted genes of *Escherichia* phage vB_EcoM_C2-3 were highly similar to those described from the *Tequatrovirus* genus, marked differences in amino acid composition were observed in some genes encoding DNA replication and transcription proteins (Fig. 3). It is worth noting that four genes were predicted to be involved in the lysis of host bacteria. Among them, an endolysin which degrades cell wall peptidoglycan, an holin that permeabilizes the cell membrane, and two spanins that contributes to breaking the outer membrane (Catalão et al., 2013). An antiholin gene was also detected, which contributes to the timing of host lysis by inhibiting the holin (Tran et al., 2007).

Taken together, our findings demonstrate that *Escherichia* phage vB_EcoM_C2-3 is a new member of T4-like phages isolated from urban sewage. Although additional studies are needed, the genomic features provide valuable insights into the use of a potential biocontrol agent against multi-drug resistant *E. coli* strains.

Author credit statement

Ana C. Maganha de Almeida Kumlien: methodology, investigation, formal analysis; **Clara Pérez-Vega**: methodology, investigation, formal analysis; **Edgar González-Villalobos**: methodology, investigation, formal analysis; **Carles M. Borrego**: conceptualization, methodology, investigation, formal analysis, writing review and editing, supervision; **José Luis Balcázar**: conceptualization, methodology, investigation, formal analysis, writing review and editing, supervision, funding acquisition.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 792686".This work was also supported by the Generalitat de Catalunya (ICRA-ENV 2017 SGR 1124), the CERCA Program of the Catalan Government, and by a New Lecturer Research Grant from the Society for Applied Microbiology. The authors thank Dr. Kaczorowski for providing us with the reference E. coli strain. EGV thanks the National Council for Science and Technology (CONACYT, Mexico) for the postdoctoral fellowship (CVU no. 441304).

References

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44, W16–W21.

Bai, A.D., Bonares, M.J., Thrall, S., Bell, C.M., Morris, A.M., 2020. Presence of urinary symptoms in bacteremic urinary tract infection: a retrospective cohort study of *Escherichia coli* bacteremia. BMC Infect. Dis. 20, 781.

Bailly-Bechet, M., Vergassola, M., Rocha, E., 2007. Causes for the intriguing presence of tRNAs in phages. Genome Res. 17, 1486–1495.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30, 2114–2120.

Bunduki, G.K., Heinz, E., Phiri, V.S., Noah, P., Feasey, N., Musaya, J., 2021. Virulence factors and antimicrobial resistance of uropathogenic *Escherichia coli* (UPEC) isolated from urinary tract infections: a systematic review and meta-analysis. BMC Infect. Dis. 21, 753.

Catalão, M. J., Gil, F., Moniz-Pereira, J., São-José, C., Pimentel, M., 2013. Diversity in bacterial lysis systems: bacteriophages show the way. FEMS Microbiol. Rev. 37, 554–571.

Chegini, Z., Khoshbayan, A., Vesal, S., Moradabadi, A., Hashemi, A., Shariati, A., 2021. Bacteriophage therapy for inhibition of multi drug-resistant uropathogenic bacteria: a narrative review. Ann. Clin. Microbiol. Antimicrob. 20, 30.

CLSI, 2016. Performance Standards for Antimicrobial Susceptibility Testing. 26th ed. CLSI supplement M100S. Clinical and Laboratory Standards Institute, Wayne, PA.

Garneau, J.R., 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.

González-Villalobos, E., Ribas-Aparicio, R.M., Montealegre, G., Belmont-Monroy, L., Ortega-García, Y., Aparicio-Ozores, G., Balcázar, J.L., Eslava-Campos, C.A., Hernández-Chiñas, U., Molina-López, J., 2021. Isolation and characterization of novel bacteriophages as a potential therapeutic option for *Escherichia coli* urinary tract infections. Appl. Microbiol. Biotechnol. 105, 5617–5629.

Kropinski, A.M., Mazzocco, A., Waddell, T.E., Lingohr, E., Johnson, R.P., 2009. Enumeration of bacteriophages by double agar overlay plaque assay. Methods Mol. Biol. 501, 69–76.

Lowe, T.M., Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44, W54–W57.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Rüger, W., 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67, 86–156.

Petkau, A., Stuart-Edwards, M., Stothard, P., Van Domselaar, G., 2010. Interactive microbial genome visualization with GView. Bioinformatics 26, 3125–3126.

Sambrook, J., Russell, D.W., 2006. Purification of nucleic acids by extraction with phenol:chloroform. Cold Spring Harb. Protoc. 1, pdb.prot4455.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.

Sullivan, M.J., Petty, N.K., Beatson, S.A., 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27, 1009–1010.

Tran, T.A., Struck, D.K., Young, R., 2007. The T4 RI antiholin has an N-terminal signal anchor release domain that targets it for degradation by DegP. J. Bacteriol. 189, 7618–7625.

Wang, R., Xing, S., Zhao, F., Li, P., Mi, Z., Shi, T., Liu, H., Tong, Y., 2018. Characterization and genome analysis of novel phage vB_EfaP_IME195 infecting *Enterococcus faecalis*. Virus Genes 54, 804–811.

van Hoffen, E., Mercenier, A., Vidal, K., Benyacoub, J., Schloesser, J., Kardinaal, A., Lucas-van de Bos, E., van Alen, I., Roggero, I., Duintjer, K., Berendts, A., Albers, R., Kleerebezem, M., ten Bruggencate, S., 2021. Characterization of the pathophysiological determinants of diarrheagenic *Escherichia coli* infection using a challenge model in healthy adults. Sci. Rep. 11, 6060.

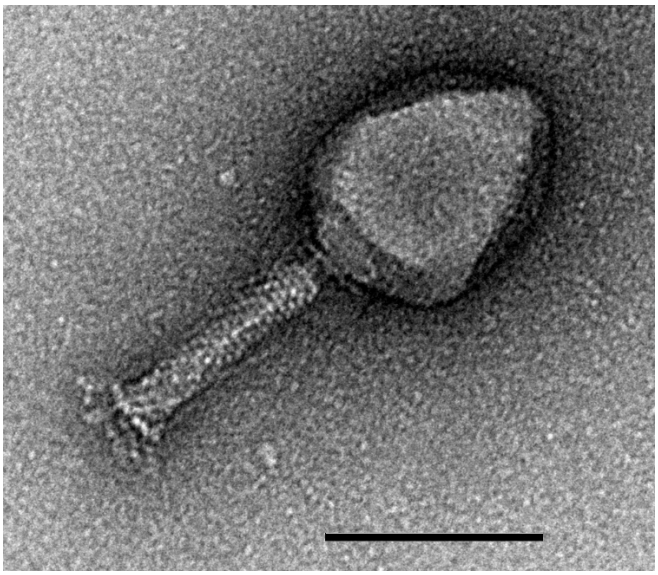
Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644.

Figure captions

Fig. 1. Transmission electron micrograph of *Escherichia* phage vB_EcoM_C2-3. The scale bar represents 100 nm.

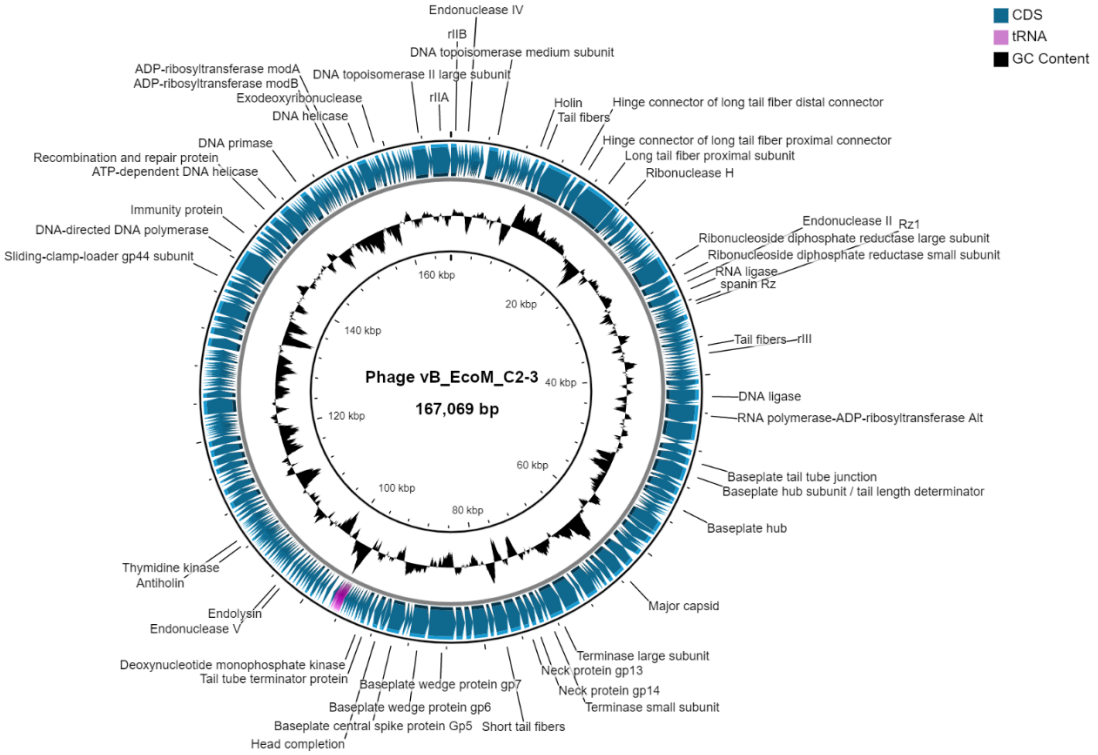
Fig. 2. Genomic map of *Escherichia* phage vB_EcoM_C2-3. Predicted coding regions are shown by arrows indicating the direction of transcription. The black circle represents the G+C content.

Fig. 3. Comparative genomic maps of *Escherichia* phage vB_EcoM_C2-3 and the closely related phages. Colored arrows indicate open reading frames (ORFs) according to their predicted function. Homologous genomic regions between phages are indicated by purple shading.

224 

225 **Figure 1**

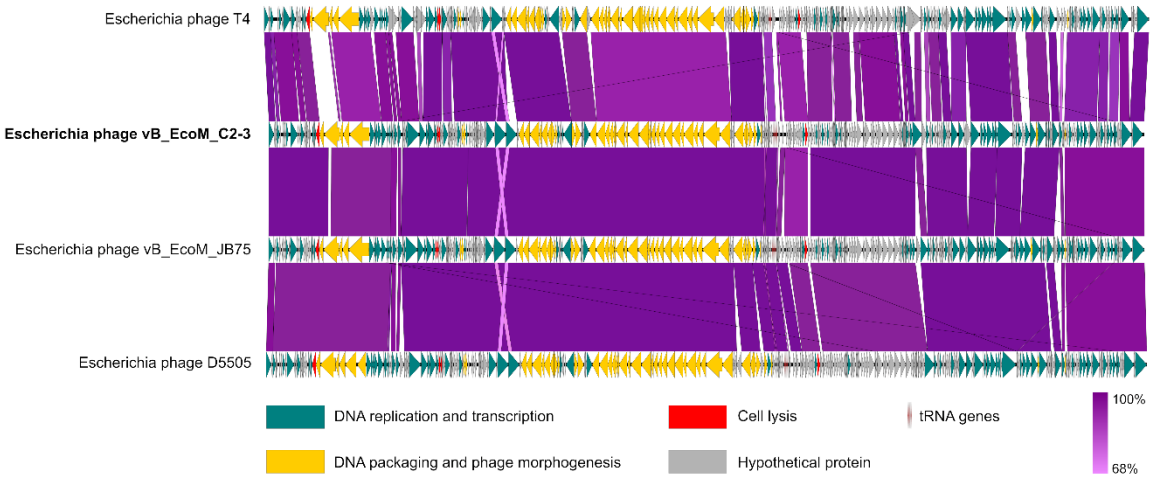
226



227

228 **Figure 2**

229



230

231 **Figure 3.**