Characterization and complete genome analysis of a novel Escherichia phage, vB_EcoM-RPN242

Napakhwan Imklin¹ · Pattaraporn Sriprasong¹ · Narut Thanantong² · Porntippa Lekcharoensuk³,4 · Rujikan Nasanit¹

Received: 27 December 2021 / Accepted: 4 April 2022 / Published online: 22 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

The novel Escherichia phage vB_EcoM-RPN242 was isolated using a strain of Escherichia coli originating from a diarrheic piglet as a host. The phage was able to form plaques on the E. coli lawn at 15–45 °C. Moreover, it was stable over a wide pH (4–10) and temperature (4–70 °C) range. The vB_EcoM-RPN242 genome was found to be a linear, double-stranded DNA consisting of 154,840 base pairs. There were 195 protein-encoding genes and two tRNAs detected in the genome; however, no genes associated with virulence, toxins or antimicrobial resistance were found. According to overall nucleotide sequence comparisons, vB_EcoM-RPN242 possibly represents a new species in the genus Agtrevirus.

A common Gram-negative bacterium, Escherichia coli, is one of the most serious problems for the swine industry. Depending on the pathotype, it can cause severe diseases such as edema disease and neonatal diarrhea. Furthermore, antibiotic resistance has become a major concern [1]. Bacterial infection has a direct detrimental impact on swine production, including product loss, increased expenses, and health concerns. As a result, medical solutions for both prevention and therapy must be developed in order to address the issue.

Handling Editor: T. K. Frey.

Porntippa Lekcharoensuk
fveptn@ku.ac.th

Rujikan Nasanit
nasanit_r@su.ac.th

1 Department of Biotechnology, Faculty of Engineering and Industrial Technology, Silpakorn University, Sanamechandra Palace Campus, Nakhon Pathom 73000, Thailand

2 Department of Farm Resources and Production Medicine, Faculty of Veterinary Medicine, Kasetsart University, Kamphaeng Saen, Nakhon Pathom, Thailand

3 Department of Microbiology and Immunology, Faculty of Veterinary Medicine, Kasetsart University, Bangkok 10900, Thailand

4 Center for Advanced Studies in Agriculture and Food, Kasetsart University Institute for Advanced Studies, Kasetsart University, Bangkok 10900, Thailand

Bacteriophages (phages) are naturally ubiquitous and are believed to be the most diverse natural materials in the biosphere [2]. Generally, they are classified according to their life cycles, which are chronic, lytic, and lysogenic [3]. These useful viruses have been employed in various fields, such as food safety, therapeutic applications, and biological studies. Focusing on therapeutic purposes, lytic phages are preferable to lysogenic phages because they cause bacterial death directly. In contrast, lysogenic phages integrate their genetic elements into the host genome and, in some cases, do not immediately kill the bacteria [4–6].

In this study, a novel virulent phage, vB_EcoM-RPN242, was isolated using E. coli M242 as a host. The host was obtained from the intestine of a diarrheic piglet and provided by Kamphaengsaen Veterinary Diagnostic Center (KVDC), Kasetsart University, Nakhon Pathom, Thailand. The phage was purified and propagated to investigate its morphology, efficiency of plating (EOP), thermal and pH stability, adsorption rate, single-step growth, and genome characteristics.

To observe phage morphology, a purified phage suspension was deposited onto a formvar-coated copper grid, stained with 2% uranyl acetate, and examined using a Hitachi HT7700 transmission electron microscope. Micrographs of the vB_EcoM-RPN242 particles revealed icosahedral heads, approximately 87 ± 5 nm in diameter, and long-contractile tails, 133 ± 6 nm in length (Fig. 1).

To obtain information about phage characteristics, several experiments were conducted. An EOP test was performed...
using the agar overlay technique. The plates were incubated at different temperatures as described by Seeley and Primrose [7]. The results indicated that the vB_EcoM-RPN242 was able to produce clear plaques on E. coli M242 at 15–45 °C. The optimal temperature for plating was 20 °C (Supplementary Fig. S1). This suggests that the phage is a mid-temperature phage [7]. For thermal stability, the phage was tested at 4, 28, 37, 50, 60, and 70 °C for 1 h in SM buffer. The pH stability of the phage was tested at various pH values (2–11) in tryptic soy broth (TSB) at 37 °C for 1 h. The results showed that this phage is stable at 4–70 °C and over a wide pH range (4–10) (Supplementary Fig. S2a and b). It was unable to survive at pH values below 4 and above 10. An adsorption test was carried out following the procedure of Merabishvili and colleagues, with some modifications [8]. The phage and the host were mixed in TSB at an MOI of 0.001, and a 400-µL aliquot was collected every 3 min for 30 min. As a control, a sample was collected from the phage-TSB mixture at the end (30 min) of the test. All samples were treated with chloroform and used for phage counting. A single-step growth experiment was conducted as described by Imklin and Nasanit [9]. The results demonstrated that approximately 97% of the phage particles adsorbed to the host cells within 3 min. At 15 min, 99% of them had attached to the cells, and cell lysis began at 21–24 min (Supplementary Fig. S3). The latent and burst periods were about 25 and 50 min, respectively. The burst size was 1074 ± 27 PFU/infected cell (Supplementary Fig. S4). These biological characteristics imply that this phage is able to lyse its host over a wide range of temperatures and survives under environmental conditions that are favorable for therapeutic purposes [10].

To analyze the phage genome, the phage genomic DNA was extracted by the phenol-chloroform procedure as described previously [9]. After a quality control step, the genomic DNA was used for DNA library construction. Sequencing was performed using an Illumina HiSeq 2500 sequencing system (Macrogen, Korea). Downstream processes were carried out to validate the raw reads, which underwent further processing using FastQC (v0.11.5) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and Trimmomatic (v0.36) [11] to obtain qualified reads. De novo assembly of the filtered reads was performed using SPAdes [12] to produce a single contig corresponding to the vB_EcoM-RPN242 phage genome. These procedures were conducted by the Macrogen bioinformatics team.

PCR was used to verify the nucleotide sequences at both ends of the contig. The PCR reactions were performed using the specifically designed primers RPN242-F (5′-GGTGATGTTTAGTGACGCAGC-3′) and RPN242-R (5′-GGT AGATGACGACGT-3′). Open reading frames (ORFs) were predicted using Glimmer 3 [13] and GeneMarkS [14]. The translated sequences were analyzed using BLASTp [15], the Conserved Domain Database [16], and HHpred [17] to predict their functions. tRNA genes in the phage genome were identified using Aragorn (v1.2.41) [18]. Phylogenetic trees based on amino acid sequence alignments were constructed in MEGA 7 using the neighbor-joining (NJ) method with 1000 bootstrap replicates [19]. The whole genome sequence of phage vB_EcoM-RPN242 was subjected to PAirwise Sequence Comparison (PASC) analysis [20] to evaluate its overall nucleotide sequence similarity to related viruses. The most closely related phage identified by PASC was selected to perform a linear comparison against the vB_EcoM-RPN242 genome and generate a visual representation of the genome using Easyfig [21].

Phage vB_EcoM-RPN242 has a linear, double-stranded DNA with a length of 154,840 bp and GC content of 48.8%. In the genome, 195 ORFs and two tRNA genes were identified; 129 ORFs were on one strand, and the remaining 66 ORFs and tRNAs genes were on the opposite strand (Fig. 2). A total of 130 genes were predicted to encode hypothetical proteins, while 65 genes encoded gene products of known function. The 65 ORFs were assigned to five modules based on the biological functions predicted by bioinformatics analysis. The group of DNA-replication-, recombination-, repair-, and packaging-associated genes included 30 ORFs, encoding HNH endonuclease, DNA topoisomerase, DNA helicase, DNA ligase, RecA-like recombination protein, ssDNA binding protein, crossover junction endodeoxyribonuclease, DNA primase, endonuclease subunit, RNase H, clamp loader, sliding clamp, nuclease, DNA repair/recombination...
Complete genome analysis of *Escherichia* phage vB_EcoM-RPN242

1677

Another 14 genes encoding proteins associated with transcription, translation, and nucleotide metabolism were predicted: deoxycytidylate deaminase, 2-deoxyuridine 5-triphosphate nucleotidohydrolase, thymidylate synthase, phage late-transcription coactivator, glutaredoxin, ribonucleoside-diphosphate reductase, PhoH-like protein, RNA polymerase sigma factor, nucleoside triphosphate pyrophosphohydrolase, translation repressor protein, 5'(3')-deoxyribonucleotidase, P-loop nucleotide kinase 2, and alpha-glutamyl/putrescinyl thymine pyrophosphorylase clade 2. Twelve ORFs were predicted to encode structural proteins, including baseplate wedge subunits, tail fiber proteins, tail spike protein, neck proteins, tail sheath protein, tail tube protein, major capsid protein, and baseplate protein. Six morphogenesis-related genes were predicted, including tail sheath stabilizer, portal vertex protein, prohead core scaffolding protein and protease, prohead core protein, tail completion and sheath stabilizer protein, and head completion protein. In addition, N-acetylmuramidase was identified as a lysozyme [22]. The baseplate hub subunit and tail lysozyme function as structural proteins and in lysozyme activity [23]. The DNA end protector protein, gp2, has been shown to be involved in host-virus interactions and to prevent DNA degradation by exonuclease V [24]. Therefore, these three genes were classified as genes associated with phage-host interaction and lysis activity.

The BLAST search results for these proteins are presented in Supplementary Table S1. No genes associated with virulence, antimicrobial resistance, or toxins, which would make them unsatisfactory for therapeutic purposes [25], were identified in the vB_EcoM-RPN242 genome. This, together with the absence of lysogeny-associated genes, suggests that this phage might be applicable for therapeutic use in veterinary medicine.

To investigate the evolutionary relationship between vB_EcoM-RPN242 and other phages, phylogenetic trees were constructed based on an alignment of the amino acid sequences of the terminase large subunit (TerL) and major capsid protein (MCP). The other phage TerL and MCP sequences were retrieved from public databases based on BLASTp homology search results. The TerL and MCP trees showed the closest relative of vB_EcoM-RPN242 to be *Escherichia* phage vB_EcoM-ZQ1 (Supplementary Fig. S5a and b). PASC analysis showed approximately 87% overall nucleotide sequence identity to vB_EcoM-ZQ1, while the other nine closest relatives showed 79–73% sequence identity (Supplementary Table S2). A visualization of the BLASTn-based genome comparison demonstrated a reciprocal genealogical relationship between phages vB_EcoM-RPN242 and vB_EcoM-ZQ1, with 65–100% nucleotide sequence identity (Fig. 3). According to the National Center for Biotechnology Information (NCBI) database, phage vB_EcoM-ZQ1 is a member of the subfamily Aglimvirinae that has not been assigned to a genus. This implies that vB_EcoM-RPN242 also cannot be ranked at the genus level. Nevertheless, when compared to a member of the genus Agtreivirus, phage EspM4VN4, the two viruses appear to have the same tail morphology [26]. Additionally, the tail spike protein gene cluster was found to be arranged in the same pattern as those of members of the family Ackermanviridae [27], starting with the baseplate wedge subunit (gp166), followed by two putative proteins (gp165, gp164), four tail spike proteins (gp163, gp162, gp161, gp160), and a virulence-associated protein (gp159). According to the
cutoff criteria for genera established by the International Committee on Taxonomy of Viruses (ICTV) Bacterial and Archaeal Viruses Subcommittee, phages are recognized as belonging to the same genus if their whole-genome nucleotide sequences share more than 70% identity [28]. Therefore, vB_EcoM-RPN242 was classified as a member of the genus *Agtrivirus*, together with the phages mentioned above that were identified as relatives in PASC analysis. Nevertheless, it was not assigned to any species in genus *Agtrivirus* recognized by ICTV, since it is less than 95% identical to any other phage. This suggests that it should be considered a member of a new species.

To evaluate whether phages are suitable for therapeutic purposes and food safety, they must be genetically characterized [29]. Study of their biological properties provides information that can be used to predict their efficacy. In conclusion, vB_EcoM-RPN242 has the ability to infect *E. coli* associated with diarrheal piglets. It is stable under a range of pH and temperature conditions and lacks any undesirable genes. Thus, phage vB_EcoM-RPN242 has potential as a biotherapeutic agent.

**Nucleotide sequence accession number** The complete genome sequence of *Escherichia* phage vB_EcoM-RPN242 is available in the GenBank database under accession number OL656110.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00705-022-05479-7.

**Acknowledgements** Laurence Crouch is credited with grammar and language proofreading.

**Author contributions** The study conception, design and supervision were contributed by Rujikan Nasanit. Material preparation was performed by Napakhwan Imklin, Pattaraporn Sriprasong, and Narut Thanantong. Data collection and analysis were performed by Napakhwan Imklin and Pattaraporn Sriprasong. Funding acquisition and research program management were contributed by Porntippa Lekcharsuk. The first draft of the manuscript was written by Porntippa Lekcharsuk. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Fig. 3** Comparison of the genome sequences of vB_EcoM-RPN242 and *Escherichia* phage vB_EcoM-ZQ1 (MW650886). The dark-light gray and green-yellow gradients indicate the similarity of the nucleotide sequences in the forward and reverse alignment, respectively. The colored arrows indicate the categories of the protein coding regions based on their putative functions.

**Funding** This work was supported by Kasetsart University Research and Development Institute, KURDI grant no. FF(KU)17.64.

**Declarations**

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**References**

1. Luppi A (2017) Swine enteric colibacillosis: diagnosis, therapy and antimicrobial resistance. Porcine Health Manag 3:16
2. Dion MB, Oechslin F, Moineau S (2020) Phage diversity, genomics and phylogeny. Nat Rev Microbiol 18:125–138
3. Chevallereau A, Pons BJ, van Houle S, Westra ER (2022) Interactions between bacterial and phage communities in natural environments. Nat Rev Microbiol 20:49–62
4. Sohail HA, Coffey A, Debowska K, Meyer IM, Middelhoe M, Sohail M, Clokie MRJ (2020) Bacteriophages: emerging applications in medicine, food, and biotechnology. Phage 1:75–82
5. Rohde C, Wittmann J (2020) Phage diversity for research and application. Antibiotics (Basel) 9:734
6. Principi N, Silvestri E, Esposito S (2019) Advantages and limitations of bacteriophages for the treatment of bacterial infections. Front Pharmacol 10:513
7. Seeley ND, Primrose SB (1980) The effect of temperature on the ecology of aquatic bacteriophages. J Gen Virol 46:87–95
8. Merabishvili M, Vandenhove D, Kropinski AM, Mast J, De Vos D, Verbeken G, Noben JP, Lavigne R, Vanechoutte M, Pirnay JP (2014) Characterization of newly isolated lytic bacteriophages active against *Acinetobacter baumannii*. PLoS One 9:e104853
9. Imklin N, Nasanit R (2020) Characterization of *Salmonella* bacteriophages and their potential use in dishwashing materials. J Appl Microbiol 129:266–277
10. Loponte R, Pagnini U, Iovane G, Pisanelli G (2021) Phage therapy in veterinary medicine. Antibiotics (Basel) 10:421
11. Bolger AM, Lohse M, Usadel B (2014) Trimмomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
12. 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
Complete genome analysis of Escherichia phage vB_EcoM-RPN242

and its applications to single-cell sequencing. J Comput Biol 19:455–477

13. Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679
14. 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
15. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008) NCBI BLAST: a better web interface. Nucleic Acids Res 36:W5–W9
16. Marchler-Bauer A, Derbyshire MK, Gonzalez NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thaiki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH (2015) CDD: NCBI’s conserved domain database. Nucleic Acids Res 43:D222–D226
17. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248
18. Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:W11–W16
19. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
20. Bao Y, Chetvernin V, Tatusova T (2014) Improvements to pairwise sequence comparison (PASC): a genome-based web tool for virus classification. Arch Virol 159:3293–3304
21. Sullivan MJ, Petty NK, Beatson SA (2011) Easyfig: a genome comparison visualizer. Bioinformatics 27:1009–1010
22. Broendum SS, Buckle AM, McGowan S (2018) Catalytic diversity and cell wall binding repeats in the phage-encoded endolysins. Mol Microbiol 110:879–896
23. Kanamaru S, Gassner NC, Ye N, Takeda S, Arisaka F (1999) The C-terminal fragment of the precursor tail lysozyme of bacteriophage T4 stays as a structural component of the baseplate after cleavage. J Bacteriol 181:2739–2744
24. Wang GR, Vianelli A, Goldberg EB (2000) Bacteriophage T4 self-assembly: in vitro reconstitution of recombinant gp2 into infectious phage. J Bacteriol 182:672–679
25. Philipson CW, Voegtly LJ, Lueder MR, Long KA, Rice GK, Frey KG, Biswas B, Cer RZ, Hamilton T, Bishop-Lilly KA (2018) Characterizing phage genomes for therapeutic applications. Viruses 10:188
26. Thanh NC, Nagayoshi Y, Fujino Y, Iiyama K, Furuya N, Hiromasa Y, Iwamoto T, Doi K (2020) Characterization and genome structure of virulent phage EspM4VN to control Enterobacter sp. M4 isolated from plant soft rot. Front Microbiol 11:885
27. Sorensen AN, Woudstra C, Sorensen MCH, Bronsted JL (2021) Subtypes of tail spike proteins predicts the host range of Ackermannviridae phages. Comput Struct Biotechnol J 19:4854–4867
28. Turner D, Kropinski AM, Adriaenssens EM (2021) A Roadmap for genome-based phage taxonomy. Viruses 13:506
29. Endersen L, O’Mahony J, Hill C, Ross RF, McAuliffe O, Coffey A (2014) Phage therapy in the food industry. Annu Rev Food Sci Technol 5:327–349

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.