



## Data Article

# Whole genome sequencing and *in vitro* activity data of *Escherichia* phage NTEC3 against multidrug-resistant Uropathogenic and extensively drug-resistant Uropathogenic *E. coli* isolates

Naveen Chaudhary<sup>a</sup>, Dharminder Singh<sup>a</sup>, Ravi Kumar Maurya<sup>a</sup>,  
Balvinder Mohan<sup>a</sup>, Ravimohan S. Mavuduru<sup>b</sup>, Neelam Taneja<sup>a,\*</sup>

<sup>a</sup> Department of Medical Microbiology, Postgraduate Institute of Medical Education and Research, Chandigarh, India

<sup>b</sup> Urology, Postgraduate Institute of Medical Education and Research, Chandigarh, India

## ARTICLE INFO

*Article history:*

Received 1 November 2021

Revised 12 July 2022

Accepted 13 July 2022

Available online 18 July 2022

*Keywords:*

Phage

Multidrug-resistant

Uropathogenic *Escherichia coli*

Urinary tract infections

Sequencing

Biofilms

## ABSTRACT

This data article describes the whole-genome sequencing and *in vitro* activity data of *Escherichia* phage NTEC3 isolated from a community sewage sample in Chandigarh, India. The phage NTEC3 was active against multi-drug-resistant (MDR) and extensively drug-resistant (XDR) biofilm-forming Uropathogenic *Escherichia coli* (UPEC) strains. The genome of this phage was linear, double-stranded, and 44.2 kb long in size. A total of 72 ORFs (open reading frames) were predicted and 30 ORFs were encoded for functional proteins. The phage belonged to the *Kagunavirus* genus of the *Siphoviridae* family. Phylogenetic analysis using DNA polymerase was performed to understand the phage evolutionary relationships. Genes encoding for lysogeny, virulence, toxins, antibiotic resistance, and the CRISPR/CRISPR-like system were not found during screening. The annotated genome was deposited in Genbank under the accession number OK539620.

\* Corresponding author.

E-mail address: [taneja.neelam@pgimer.edu.in](mailto:taneja.neelam@pgimer.edu.in) (N. Taneja).

## Specifications Table

Subject	Microbiology
Specific subject area	Medical microbiology
Type of data	WGS and <i>in vitro</i> testing data are provided in tables and figures.
How the data were acquired	WGS data were generated using Illumina Novaseq 6000 platform
Data format	Analyzed
Description of data collection	The phage NTEC3 was isolated from a community sewage water in Chandigarh, India. DNA was extracted and WGS has performed on Illumina Novaseq 6000 sequencer with a paired-end library of a read length of $2 \times 150$ bp [1,2]. The genome was assembled using the strategic k-mer extension for scrupulous assemblies (SKESA v2.4.0) assembler with default k-mer sizes. The ORFs were predicted using a gene locator and interpolated markov modeler (GLIMMER v3.02) and GeneMarkS v4.28 and subsequently annotated with PHASTER and RAST servers.
Data source location	Institution: Postgraduate Institute of Medical Education and Research City/Town/Region: Chandigarh Country: India Latitude and longitude for collected samples/data: 30.7650° N, 76.7750° E
Data accessibility	Repository name: GenBank, Sequence Reads Archive (SRA) Data identification numbers: OK539620, SRP369723 The direct URL to the data is as follows: <a href="https://www.ncbi.nlm.nih.gov/nucleotide/OK539620">https://www.ncbi.nlm.nih.gov/nucleotide/OK539620</a> <a href="https://trace.ncbi.nlm.nih.gov/Traces/index.html?view=study&amp;acc=SRP369723">https://trace.ncbi.nlm.nih.gov/Traces/index.html?view=study&amp;acc=SRP369723</a>

## Value of the Data

- Data provides genomic information on a lytic phage to researchers for sequence comparison and evolutionary relationship studies.
- Genome sequencing and *in vitro* activity data of the phage can be used by the scientific community for screening and identification of novel phage-based antimicrobial strategies.
- Data provides information about a potentially safe product for therapeutics against drug-resistant UPEC as the phage lacked genes encoding for lysogeny, virulence, toxins, and antibiotic resistance.
- Data provides information about a lytic enzyme endolysin whose sequence could be used to design recombinant endolysins to treat biofilm-associated infections.

## 1. Data Description

UPEC strains have a variety of virulence-associated factors (VFs) like adhesins, toxins, siderophores, chaperone-usher (CU) fibers, invasins, and serum resistance-associated proteins that help to invade and injure the host [3].

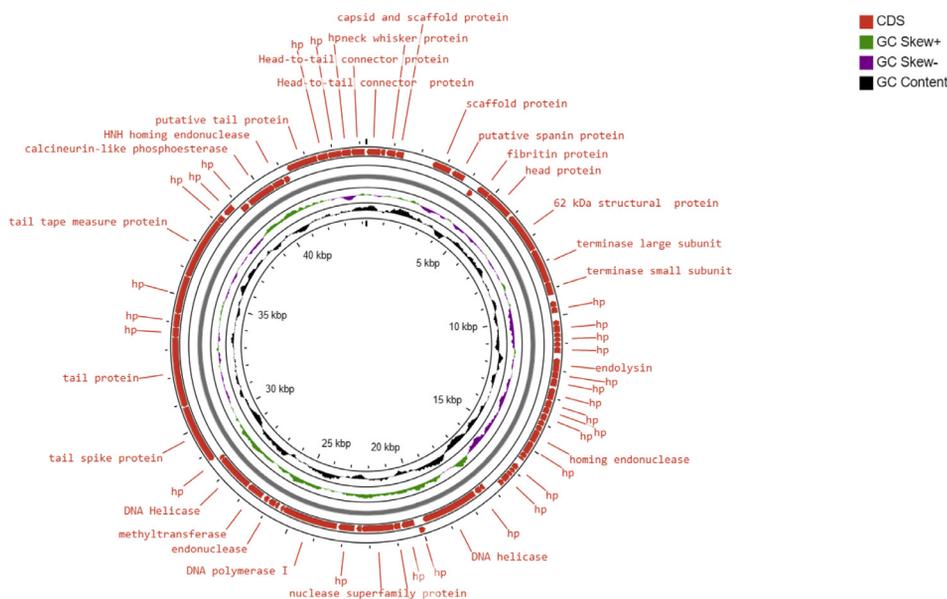
Many UTIs causing UPEC strains carry different antibiotic resistance genes like Extended-Spectrum- $\beta$ -Lactamases (ESBLs) and Metallo- $\beta$ -Lactamases (MBLs) on their chromosomes. UTIs caused by MDR and XDR UPECs account for one of the major therapeutic challenges in the health sector [4]. The widespread decline in antibiotic effectiveness has sparked renewed interest in alternative therapeutics like phage therapy. Phage therapy is primarily based on the use of obligately lytic phages to eliminate their bacterial hosts while leaving human cells unharmed.

Phage NTEC3 was isolated from community sewage water in Chandigarh, India using a clinical strain UPEC 590B as a host bacterium. The NEBNext Ultra kit was used to prepare the se-

**Table 1**

Genome sequence characteristic of phage NTEC3.

Index	Value
Raw reads generated	6974384
Read length	150 bp
Library type	Paired-end
Genome size	44240
GC content	51%
Number of predicted genes with significant BLASTX match (E-value $\leq 1e-3$ and Similarity score $\geq 40\%$ ) with uniprot	72
Accession no.	OK539620
No. of Lysis /adhesion-related proteins	2
Number of tRNAs	0
CRISPR-cas sequence	0

**Fig. 1.** Circular genome view of phage NTEC3 constructed using CGView.

quencing library, and sequencing was performed on Novaseq 6000 sequencer, yielding 6974384 paired-end (150-bp-long) raw reads (Table 1). The complete genome of the phage was 44.2 kb in length, with a gene density of 1.60/kbp, and a GC content of 51% (Table 1, Figs. 1 and 2). Out of the 72 predicted ORFs, 21 (27.3%) were found to be present in the direct strand, and the rest were found in the complementary strand (Fig. 1). Forty-four ORFs (61.6%) were predicted to encode for hypothetical proteins whereas 30 ORFs (41 %) were predicted to encode for functional proteins. The functional proteins were categorized into the following groups (i) DNA replication/metabolism-related proteins (ii) Host lysis and adhesion-related proteins (iii) DNA packaging proteins (iv) Structural proteins (Table S1). The genome of this phage lacked genes encoded for temperate phage markers, toxins, virulence factors, antibiotic resistance, or the CRISPR/CRISPR-like system. The ORF 22 of phage NTEC3 encoded for an endolysin that has an intrinsic feature to disrupt biofilm mass and bacterial cell lysis by breaking the peptidoglycan layer [5]. Phage NTEC3 formed plaques of 4-5 mm in diameter (Fig. 3). The phage was active against 24.4% of 45 MDR and XDR UPEC strains resistant to third-generation cephalosporins,

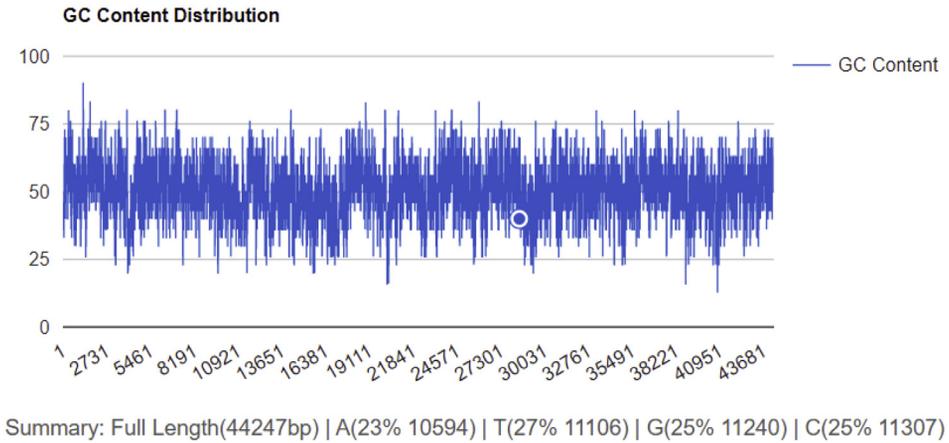


Fig. 2. GC content distribution pattern of phage NTEC3.

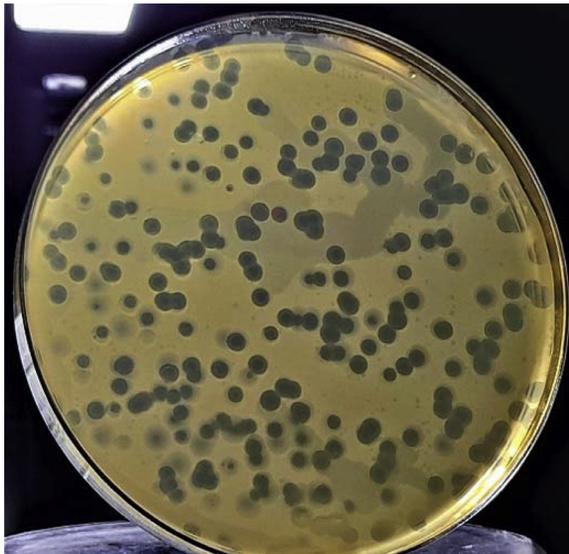
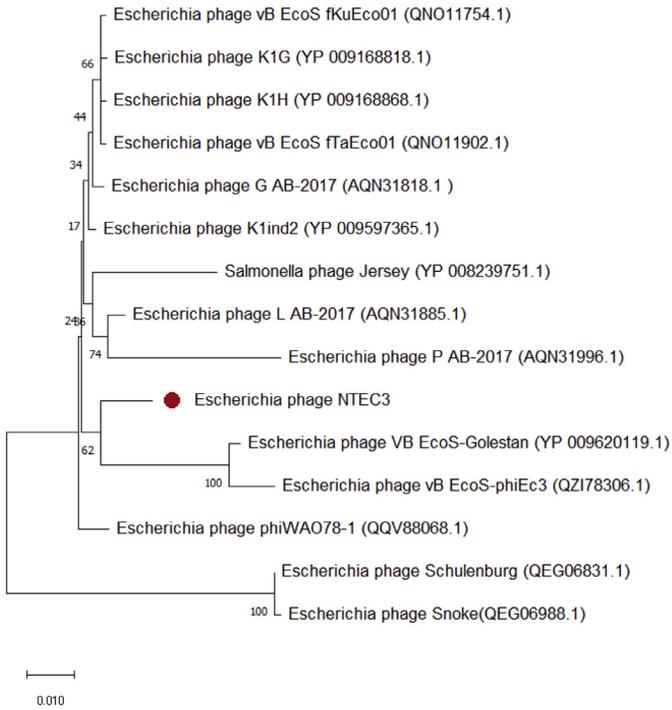


Fig. 3. Plaque morphology of phage NTEC3.

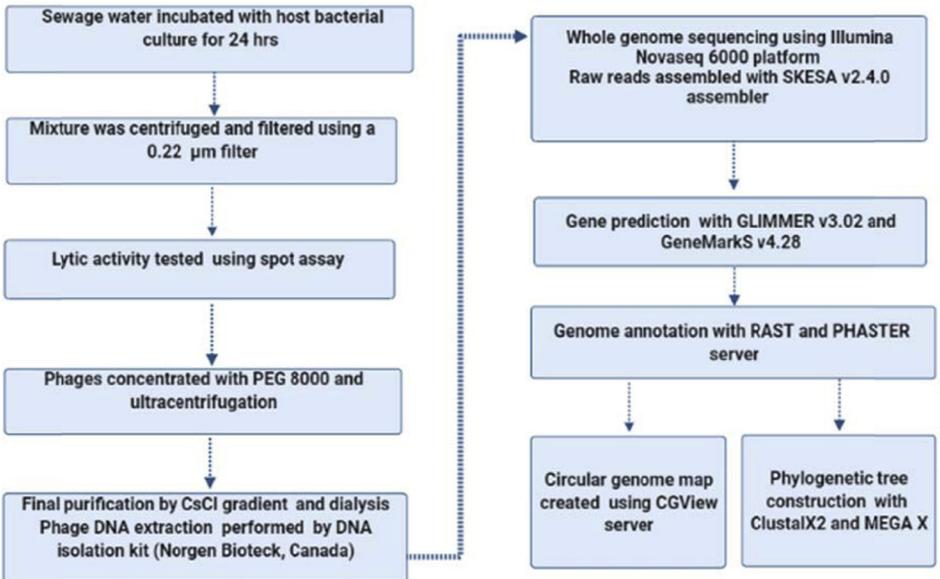
fluoroquinolones, aminoglycosides, beta-lactamase inhibitor combinations, cotrimoxazole, nitrofurantoin and imipenem (Table S2). In the phylogenetic tree, phage NTEC3 was placed in an outgroup of two *Siphoviridae* family phages *Escherichia* phage VB\_EcoS-Golestan (BLASTP identity >93.18%) and *Escherichia* phage vB\_EcoS-phiEc3 (BLASTP identity >92.78%) isolated from wastewater samples in USA and Spain, respectively (Fig. 4).

## 2. Experimental Design, Materials and Methods

A flow chart was designed using Cmap server to demonstrate the experimental design and methods (Fig. 5) [6,7].



**Fig. 4.** The neighbor-joining tree constructed based on the amino acid sequence of DNA polymerase protein using MEGA-X. Bootstraps values in the percentage of 1000 replicates are depicted next to the branches.



**Fig. 5.** Flow chart showing the experimental design protocol used for phage isolation and sequencing.

## 2.1. Phage Isolation

The *Escherichia* phage NTEC3 was isolated from the community sewage treatment plant in Chandigarh. MDR UPEC 590 strain was used as host bacterium to amplify and propagate phage NTEC3. The raw sewage water samples were centrifuged at 1500 RPM and the supernatant was filtered through a 0.45  $\mu\text{m}$  membrane filter. The filtrate was incubated at 37°C with different bacterial cultures for 18 h. After incubation, the mixture was centrifuged at 4000 RPM and filtered using a 0.22  $\mu\text{m}$  syringe filter. The lytic activity of the filtrate was evaluated using a spot assay against the respective bacterial strains.

## 2.2. Phage Purification

A clear spot on agar was scrapped and incubated with the host bacterial culture for overnight incubation at 37°C. A single isolated plaque was picked after repeating the plaque assay experiment thrice. Phage plaque size was measured and expressed in millimeters. Phages were concentrated using the overnight polyethylene glycol (PEG 8000) precipitation method followed by ultracentrifugation (50,000 RPM) and the ultrafiltration method [8]. Final purification was performed with CsCl gradient and dialyzed against SM buffer (pH 7.5) using tubing membrane (MWCO 12,000). Host range activity of phage NTEC3 was tested against 45 MDR and XDR UPEC strains using the standardized spot assay method.

## 2.3. Genome Sequencing and Analysis

The DNA extraction was performed with a phage DNA isolation kit (Norgen Bioteck, Canada). The whole genome of the phage was sequenced using the Illumina Novaseq 6000 platform with a paired-end library of a read length of  $2 \times 150$  bp [9].

SKESA v2.4.0 assembler was used to execute de novo assembly with default k-mer sizes [10]. GLIMMER v3.02 and GeneMarkS v4.28 were used to predict genes from the assembled sequence and annotated with PHASTER and RAST server and also searched against BLASTP in the UniProt database [11–13]. The CGView server was used to make a circular genomic map of the phage genome [14]. The amino acid sequences of the DNA polymerase gene of similar phages with BLASTp identity of >90% were used for constructing the phylogenetic tree.

## Ethics Statements

This study was approved by the Institute Ethical Clearance Committee of the postgraduate institute of Medical Education and Research (Chandigarh, India).

## Declaration of Competing Interest

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.

## Data Availability

[SRA accession number \(Original data\)](#) (SRA).

[GenBank accession number \(Original data\)](#) (GenBank).

## CRedit Author Statement

**Naveen Chaudhary:** Methodology, Software, Data curation, Writing – original draft; **Dharminder Singh:** Methodology, Data curation; **Ravi Kumar Maurya:** Methodology, Data curation; **Balvinder Mohan:** Supervision, Validation, Data curation; **Ravimohan S. Mavuduru:** Supervision, Validation, Data curation; **Neelam Taneja:** Supervision, Resources, Conceptualization, Validation, Writing – review & editing.

## Acknowledgments

The study work was supported by the [University Grants Commission](#) (New Delhi, India) (grant [313690](#)) and the Department of Biotechnology (New Delhi, India) (grant [102/IFD/SAN/2138/2019-2020](#)).

## Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.dib.2022.108479](#).

## References

- [1] N. Chaudhary, D. Singh, C. Narayan, B. Samui, B. Mohan, R.S. Mavuduru, N. Taneja, Complete genome sequence of Escherichia phage 590B, active against an extensively drug-resistant Uropathogenic Escherichia coli isolate, *Microbiol. Resour. Announc.* 10 (2021) 9–10, doi:[10.1128/mra.00550-21](#).
- [2] N. Chaudhary, B. Mohan, N. Taneja, Draft genome sequence of Escherichia phage PGN829.1, active against highly drug-resistant Uropathogenic Escherichia coli, *Microbiol. Resour. Announc.* 7 (2018) 20, doi:[10.1128/mra.01141-18](#).
- [3] P. Behzadi, Classical chaperone-usher (CU) adhesive fimbriome: Uropathogenic Escherichia coli (UPEC) and urinary tract infections (UTIs), *Folia Microbiol.* 65 (2020) 45–65 (Praha), doi:[10.1007/s12223-019-00719-x](#).
- [4] P. Behzadi, H.A. García-Perdomo, T.M. Karpiński, Metallo- $\beta$ -lactamases: a review, *Mol. Biol. Rep.* 47 (2020) 6281–6294.
- [5] M. Schmelcher, D.M. Donovan, M.J. Loessner, Bacteriophage endolysins as novel antimicrobials, *Future Microbiol.* 7 (2012) 1147–1171, doi:[10.2217/fmb.12.97](#).
- [6] P. Behzadi, M. Gajdács, Writing a strong scientific paper in medicine and the biomedical sciences: a checklist and recommendations for early career researchers, *Biol. Future* 72 (2021) 395–407, doi:[10.1007/s42977-021-00095-z](#).
- [7] R. Ranjbar, P. Behzadi, A. Najafi, R. Roudi, DNA microarray for rapid detection and identification of food and water borne bacteria: from dry to wet lab, *Open Microbiol. J.* 11 (2017) 330–338, doi:[10.2174/1874285801711010330](#).
- [8] V. Hietala, J. Horsma-Heikkinen, A. Carron, M. Skurnik, S. Kiljunen, The removal of endo- and enterotoxins from bacteriophage preparations, *Front. Microbiol.* 10 (2019) 1–9, doi:[10.3389/fmicb.2019.01674](#).
- [9] A. Senabouth, S. Andersen, Q. Shi, L. Shi, F. Jiang, W. Zhang, K. Wing, M. Daniszewski, S.W. Lukowski, S.S.C. Hung, Q. Nguyen, L. Fink, A. Beckhouse, A. Pébay, A.W. Hewitt, J.E. Powell, Comparative performance of the BGI and Illumina sequencing technology for single-cell RNA-sequencing, *NAR Genom. Bioinf.* 2 (2020) 1–10, doi:[10.1093/nargab/lqaa034](#).
- [10] A. Souvorov, R. Agarwala, D.J. Lipman, SKESA: strategic k-mer extension for scrupulous assemblies, *Genom. Biol.* 19 (2018) 1–13, doi:[10.1186/s13059-018-1540-z](#).
- [11] D. Arndt, J.R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, D.S. Wishart, PHASTER: a better, faster version of the PHAST phage search tool, *Nucleic Acids Res.* 44 (2016) W16–W21, doi:[10.1093/nar/gkw387](#).
- [12] R.K. Aziz, D. Bartels, A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formisano, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology, *BMC Genom.* 9 (2008) 1–15, doi:[10.1186/1471-2164-9-75](#).
- [13] A.L. Delcher, K.A. Bratke, E.C. Powers, S.L. Salzberg, Identifying bacterial genes and endosymbiont DNA with glimmer, *Bioinformatics* 23 (2007) 673–679, doi:[10.1093/bioinformatics/btm009](#).
- [14] J.R. Grant, P. Stothard, The CGView server: a comparative genomics tool for circular genomes, *Nucleic Acids Res.* 36 (2008) W181–W184, doi:[10.1093/nar/gkn179](#).