

Complete Genome Sequences of Four Novel *Escherichia coli* Bacteriophages Belonging to New Phage Groups

Alexander B. Carstens,^{a,b} Witold Kot,^{a,b} Lars H. Hansen^a

Department of Environmental Science, Aarhus University, Roskilde, Denmark^a; Department of Biology, Faculty of Science, University of Copenhagen, Copenhagen, Denmark^b

Here, we describe the sequencing and genome annotations of a set of four *Escherichia coli* bacteriophages (phages) belonging to newly discovered groups previously consisting of only a single phage and thus expand our knowledge of these phage groups.

Received 29 May 2015 Accepted 15 June 2015 Published 16 July 2015

Citation Carstens AB, Kot W, Hansen LH. 2015. Complete genome sequences of four novel *Escherichia coli* bacteriophages belonging to new phage groups. *Genome Announc* 3(4):e00741-15. doi:10.1128/genomeA.00741-15.

Copyright © 2015 Carstens et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Lars H. Hansen, lhha@envs.au.dk.

Since Félix d’Hérelle’s discovery of the bacteriophage in 1917 (1), much effort has been put into understanding phages and their diversity (2, 3). This is especially true for phages of classic laboratory strains of *Escherichia coli* (2). Nonetheless, even with almost a hundred years of studies, the diversity of *E. coli* phages is not fully explored (3). Here, we present the complete sequences of four novel *E. coli* phages belonging to two newly discovered bacteriophage groups. This will extend our knowledge of the sequence diversity of *E. coli* phages.

The four phages were isolated from animal fecal samples on *E. coli* strain MG1655. The DNA isolation and library preparation was done using direct plaque sequencing as described before (4), with the minor modification that SDS was used in a final concentration of 0.1%. The library was sequenced using Illumina paired-end sequencing on the MiSeq platform as a part of a flowcell (250 cycles; Illumina). The total read yield was between 49,884 (CAjan) and 182,112 (JenP2) reads per library. Reads were trimmed and *de novo* assembled using CLC Genomic Workbench version 7.0.4 (CLC bio, Aarhus, Denmark). The assembly was cross-verified by SPAdes Genome Assembler version 3.1.0 (5).

Three regions with relatively low coverage in CAjan’s genome were confirmed by PCR, followed by Sanger sequencing reactions (Macrogen, Seoul, Korea). Automated annotation and open reading frame identification was performed using the RAST annotation server (6). This was followed by manual verification by screening all the predicted proteins against the NCBI protein database using BLASTp (7) and by using a Pfam domain search (8). CAjan, JenP1, JenP2 and JenK1 are all double-stranded DNA phages with genome sizes of 59,670, 607,454, 59,802, and 60,747 bp, respectively. The four phages have a G/C content between 43.2% (JenP2) and 44.7% (CAjan) with an average coverage in the assembly between 114.7-fold (CAjan) and 542.79-fold (JenP2). The genomes contain between 88 (JenK1) and 91 (CAjan) predicted genes, the majority of which could not be assigned a function. All four phages contain an operon with genes involved in queuosine biosynthesis, as has been seen in other phages (9). The function of queuosine is not completely understood but might be involved in the specificity of transcription and

codon recognition (10) and might help the phage hijack the host transcription and translation machinery.

JenP1, JenP2, and JenK1 are closely related to enterobacterium phage 9g (GenBank accession number NC_024146), sharing 93%, 90%, and 96% sequence identity, respectively. Enterobacterium phage 9g was, until now, considered the sole representative of a novel phage genus (11). At the time of assembly CAjan had no close relatives, but a new *Escherichia* phage “Seurat” sharing 96% sequence identity with CAjan has recently been published (12). Seurat also belongs to a novel group of phages with no other close relatives. The four sequenced phages JenP1, JenP2, JenK1, and CAjan therefore expand the two phage groups and bring new insights to the diversity of *E. coli* phages.

Nucleotide sequence accession numbers. Genome sequences have been deposited in GenBank under the accession numbers KP064094 (CAjan), KP719132 (JenP1), KP719133 (JenP2), and KP719134 (JenK1).

ACKNOWLEDGMENTS

This work was supported by Lundbeck grant R44-A4384 and a PHAGEMID project financed by the Danish Research Council for Technology and Production (4093-00198).

REFERENCES

- D’Herelle F. 1917. On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr F D’Herelle, presented by Mr Roux. 1917. *Res Microbiol* 158:553–554.
- Campbell A. 2003. The future of bacteriophage biology. *Nat Rev Genet* 4:471–477. <http://dx.doi.org/10.1038/nrg1089>.
- Ackermann HW. 2011. Bacteriophage taxonomy. *Microbiol Aust* 32: 90–94.
- Kot W, Vogensen FK, Sørensen SJ, Hansen LH. 2014. DPS—a rapid method for genome sequencing of DNA-containing bacteriophages directly from a single plaque. *J Virol Methods* 196:152–156. <http://dx.doi.org/10.1016/j.jviromet.2013.10.040>.
- 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 and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.

6. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
7. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.
8. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301. <http://dx.doi.org/10.1093/nar/gkr1065>.
9. Sabri M, Häuser R, Ouellette M, Liu J, Dehbi M, Moeck G, García E, Titz B, Uetz P, Moineau S. 2011. Genome annotation and intraviral interactome for the *Streptococcus pneumoniae* virulent phage Dp-1. *J Bacteriol* 193:551–562. <http://dx.doi.org/10.1128/JB.01117-10>.
10. Vinayak M, Pathak C. 2010. Queuosine modification of tRNA: its divergent role in cellular machinery. *Biosci Rep* 30:135–148.
11. Kulikov EE, Golomidova AK, Letarova MA, Kostryukova ES, Zelenin AS, Prokhorov NS, Letarov AV. 2014. Genomic sequencing and biological characteristics of a novel *Escherichia coli* bacteriophage 9g, a putative representative of a new *Siphoviridae* genus. *Viruses* 6:5077–5092. <http://dx.doi.org/10.3390/v6125077>.
12. Doan DP, Lessor LE, Hernandez AC, Kutty Everett GF. 2015. Complete genome sequence of enterotoxigenic *Escherichia coli* siphophage Seurat. *Genome Announc* 3(1):e00044-15. <http://dx.doi.org/10.1128/genomeA.00044-15>.