 doi:10.20944/preprints201807.0156.v1

# The Complete Genome Sequence of *Escherichia coli* phage Eco_BIFF

Lauren Cooper [1], Erica Lasek-Nesselquist[2], and Joseph T. Wade[1,2*]

[1]Department of Biomedical Sciences, School of Public Health, University at Albany, SUNY, Albany, New York, USA.

[2]Wadsworth Center, New York State Department of Health, Albany, New York, USA.

*Corresponding author:

joseph.wade@health.ny.gov

Keywords: Bacteriophage; *Escherichia coli*; Genome sequence.

 doi:10.20944/preprints201807.0156.v1


# Abstract

*Escherichia coli* phage Eco_BIFF was isolated from several laboratory stocks of *E. coli* K-12 MG1655 derivatives. The source of the contamination is unknown. Eco_BIFF is a lytic phage that shows effective growth inhibition of *E. coli* K-12. Here, we announce the complete genome sequence of Eco_BIFF, and major findings from its genome annotation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0156.v1

# Genome Announcement

*Escherichia coli* K-12 MG1655 is a common laboratory strain used in molecular biology research both as a tool and as a model organism (Blattner et al., 1997). Eco_BIFF was isolated from strain AMD536, a previously described derivative of *Escherichia coli* K-12 strain MG1655 (Cooper et al., 2018). All AMD536 strains infected with Eco_BIFF contained two plasmids: a pBAD30 (Guzman et al., 1995) derivative and a pPRO18 (Lee and Keasling, 2005) derivative. Phage infection was suspected when cells grown in LB at 37°C with aeration lysed in early exponential phase.

*Escherichia* phage Eco_BIFF DNA was isolated from cell lysates of contaminated bacterial cultures using phenol:chloroform:isoamyl alcohol / Sodium dodecyl sulfate (PCI/SDS) DNA extraction (Phagehunting Protocols, http://phagesdb.org/phagehunters/). A library for DNA sequencing was prepared using the Nextera kit (Illumina). Whole genome sequencing was performed using an Illumina MiSeq Instrument (Wadsworth Center Applied Genomic Technologies Core).

Data generated from the genomic library yielded 262,824 paired-end reads with a read length of 2 x 251 bp. To exclude bacteria sequences, reads were first mapped to the *E. coli* K-12 MG1655 genome reference sequence (U00096.3) using CLC Genomics Workbench. All remaining sequence reads were assembled into a single 49,372-bp contig corresponding to the whole genome of Eco_BIFF. Annotation and comparative analysis were performed using PHASTER (Arndt et al., 2016; Zhou et al., 2011). Rfam was used to search for non-coding RNA genes, structured *cis*-regulatory elements, and self-splicing RNAs (Kalvari et al., 2018; Nawrocki et al., 2015). The

Eco_BIFF genome has (i) a GC content of 45.26%, (ii) 76 predicted coding regions, and (iii) no predicted non-coding RNA genes, structured *cis*-regulatory elements or self-splicing RNAs.

The *Escherichia* phage Eco_BIFF genome contains functional genes related to phage architecture and packaging machinery (head protein, portal protein, terminase protein), tail structure for host interaction (tail proteins, tail assembly protein, tail fiber proteins, tail tape measure protein), phage DNA synthesis (ATP-dependent helicase, DNA primase, methylases), and host lysis and degradation (endolysin, holin, exodeoxyribonuclease, polynucleotide kinase/phosphatase, HNH endonucleases and recombination protein). The 23 predicted genes with an assigned function, which are all homologues of genes from related phages, are scattered over the genome and are interspaced with 53 hypothetical proteins, 52 of which are homologous to proteins predicted in related bacteriophage. One hypothetical protein, however, does not share homology to related phages and, therefore, may be an annotation artifact. The most closely related phages to *Escherichia* phage Eco_BIFF include: *Escherichia* phage vB_EcoS_SH2 (KY985004.1), *Escherichia* phage ADB-2 (JX912252.1, Bhensdadia et al., 2013), *Escherichia* phage JMPW2 (KU194205.1), *Shigella* phage SH6 (KX828710.1), and *Enterobacteria* phage T1 (AY216660.1, Roberts et al., 2004). These related phages share >94% nucleotide sequence identity with >86% query coverage, and BLAST E-values of 0 (Search parameters: blastn program, Word Size: 28, expected value: 10, Hitlist size: 100, Match/Mismatch scores: 1, -2, Gapcosts: 0, 2.5, Low Complexity Filter, Filter string: L;m;: Zhang et al., 2000).

 doi:10.20944/preprints201807.0156.v1

73 **Nucleotide Sequence Accession Number.** The complete sequence of phage Eco_BIFF can be
74 accessed under the GeneBank accession number MH285980.

 doi:10.20944/preprints201807.0156.v1


## Acknowledgements

We thank the Wadsworth Center Applied Genomic Technologies Core Facility for MiSeq sequencing. We thank the Wadsworth Center Media and Tissue Culture Core Facility, the Wadsworth Center Glassware Core Facility, and the Wadsworth Center Bioinformatics Core Facility for technical support. We thank Keith Derbyshire and Pallavi Ghosh for helpful discussions.

# References

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. *44*, W16-21.

Bhensdadia, D.V., Bhimani, H.D., Rawal, C.M., Kothari, V.V., Raval, V.H., Kothari, C.R., Patel, A.B., Bhatt, V.D., Parmar, N.R., Sajnani, M.R., et al. (2013). Complete Genome Sequence of Escherichia Phage ADB-2 Isolated from a Fecal Sample of Poultry. Genome Announc. *1*, e0004313.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453–1462.

Cooper, L.A., Stringer, A.M., and Wade, J.T. (2018). Determining the Specificity of Cascade Binding, Interference, and Primed Adaptation In Vivo in the Escherichia coli Type I-E CRISPR-Cas System. MBio *9*, e02100-17.

Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. *177*, 4121–4130.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. *46*, D335–D342.

Lee, S.K., and Keasling, J.D. (2005). A propionate-inducible expression system for enteric bacteria. Appl. Environ. Microbiol. *71*, 6856–6862.

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J., et al. (2015). Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. *43*, D130–D137.

Roberts, M.D., Martin, N.L., and Kropinski, A.M. (2004). The genome and proteome of coliphage T1. Virology *318*, 245–266.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. J. Comput. Biol. J. Comput. Mol. Cell Biol. *7*, 203–214.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011). PHAST: a fast phage search tool. Nucleic Acids Res. *39*, W347-352.