



Genome Sequences of 16 *Escherichia coli* Bacteriophages Isolated from Wastewater, Pond Water, Cow Manure, and Bird Feces

 Amira R. Vitt,^a Stephen J. Ahern,^a Michela Gambino,^a Martine C. H. Sørensen,^a  Lone Brøndsted^a

^aDepartment of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg C, Denmark

ABSTRACT *Escherichia coli* is a highly diverse bacterial species comprising both commensal and pathogenic strains. Here, we report complete genome sequences of 16 *E. coli* bacteriophages isolated from various environmental samples using the ECOR collection as isolation hosts.

Escherichia coli is a versatile and genetically diverse species inhabiting the gut of animals and humans as both commensal or pathogenic strains (1). Similarly, *E. coli* phages are highly diverse currently belonging to 11 different families and 100 different genera (2) and have been isolated from a variety of sources (3–7). Here, we report the complete genomes of 16 phages isolated from diverse environmental sources.

Phages were isolated using all of the 72 strains in the *E. coli* Reference Collection (ECOR) (8), representing the diversity of *E. coli*, as isolation hosts. Samples for phage isolation were collected from a variety of sources (i.e., wastewater, pond water, cow manure, and bird feces) in Copenhagen, Denmark, enriched overnight in the presence of ECOR hosts, serially diluted, and 3 drops of 10 μ l were spotted on the same host (Table 1). Single plaques were isolated and plaque purified for at least three rounds as described elsewhere (9). Selected phage genomic DNA was isolated from 1 mL crude phage lysate by using the standard phenol-chlorophorm extraction followed by ethanol precipitation protocol (10). Genomic DNA libraries were prepared using Nextera XT kit and sequenced using Illumina MiSeq platform (250 bp paired end). The raw reads were trimmed and used for *de novo* genome assembly with CLC Genomics Workbench 9.5.3 (Qiagen, Århus). The assembled genomes were annotated using Rapid Annotations using Subsystems Technology (RAST) (11). All tools were run with default parameters.

Phage genera were preliminary assigned by overall genome BLAST similarities to the closest related phage genome available at the NCBI and stated in Table 1. The most diverse phages were isolated from pond water and wastewater, representing four different families, with phages from more genera represented in the wastewater.

The genomes ranged in size from 44544 bp (EC120) to 171814 bp (EC128, EC150) encoding 55 to 271 predicted open reading frames (Table 1). The G+C content ranged from 35.3% (EC128) to 54.6% (EC115). Also, the number of predicted tRNAs varied between the phages (none to 20 tRNAs). The largest number of the predicted tRNAs was observed in phage EC106.

All isolated phages showed similarity to known *E. coli* phages, determined by whole-genome BLAST similarity search at the NCBI database and stated in Table 1. Phage EC167 was the least similar to other coliphages with 88% query coverage and 88.5% sequence identity to phage vB_EcoS-95 (NC_048132). Phage EC148 showed the highest sequence homology to phage T5_ev212 (99.97% sequence identity, 99% query coverage). For the remaining phages in the collection, closely related phages had between 90 and 96% sequence identity over 90 to 93% query coverage. In summary, while many of the *E. coli* phages in our collection were closely related to the known

Editor Kenneth M. Stedman, Portland State University

Copyright © 2022 Vitt et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Lone Brøndsted, lobr@sund.ku.dk.

The authors declare no conflict of interest.

Received 27 June 2022

Accepted 15 September 2022

Published 28 September 2022

TABLE 1 Properties of 16 *Escherichia coli* bacteriophage genomes

Name	Isolation host	Isolation source	Family	Genus	Genome length (bp)	Total read count	Avg coverage	Genome length (bp)	ORFs ^a	tRNA	G+C (%)	Accession	Closest relative	Coverage (%)	Identity (%)	Accession
EC120	ECOR4	Wastewater	Autographiviridae	Vectrevirus	44544	674959	3327	44544	55	ND ^b	44.9	ON185580.1	<i>Escherichia</i> phage vB_EcoS_FFH_1	90	96.09	KJ190157.1
EC100	ECOR4	Pond water	Demerecviridae	Tequintavirus	108723	222491	433	108723	155	19	39	OK665835.1	<i>Escherichia</i> phage vB_EcoS_FFH_1	90	96.09	KJ190157.1
EC104	ECOR19	Bird feces	Demerecviridae	Tequintavirus	108711	163981	382	108711	155	19	39	ON185581.1	<i>Escherichia</i> phage vB_EcoS_FFH_1	90	96.09	KJ190157.1
EC105	ECOR20	Bird feces	Demerecviridae	Tequintavirus	108732	150574	319	108732	155	19	39	ON185582.1	<i>Escherichia</i> phage vB_EcoS_FFH_1	90	96.08	KJ190157.1
EC122	ECOR20	Bird feces	Demerecviridae	Tequintavirus	108723	360008	636	108723	156	19	39	ON185583.1	<i>Escherichia</i> phage SPT5	90	92.08	NC_048627.1
EC142	ECOR4	Bird feces	Demerecviridae	Tequintavirus	108723	123228	204	108723	155	19	39	ON185584.1	<i>Escherichia</i> phage vB_EcoS_FFH_1	90	96.09	KJ190157.1
EC148	ECOR4	Bird feces	Demerecviridae	Tequintavirus	107754	71380	187	107754	157	19	39	ON185585.1	<i>Escherichia</i> phage phillLS	92	92.49	NC_047822.1
EC125	ECOR13	Wastewater	Drexelviriidae	Warwickvirus	50477	534498	2617	50477	83	ND	44.6	ON185586.1	<i>Escherichia</i> phage tuinn	97	98.72	MN850606.1
EC167	ECOR20	Wastewater	Drexelviriidae	Warwickvirus	50988	487226	1733	50988	82	ND	44.2	ON185587.1	<i>Escherichia</i> phage vB_EcoS-95	88	88.45	NC_048132.1
EC195	ECOR34	Pond water	Drexelviriidae	Warwickvirus	50477	171094	485	50477	83	ND	44.6	ON185588.1	<i>Escherichia</i> phage tuinn	97	98.76	MN850606.1
EC147	ECOR4	Wastewater	Drexelviriidae	Hannriverivirus	46060	587499	3008	46060	77	ND	43.9	ON185589.1	<i>Escherichia</i> phage egaa	95	95.98	NC_049852.1
EC101	ECOR4	Pond water	Straboviriidae	Tequatrovirus	167718	201824	270	167718	264	10	35.4	OL310488.1	<i>Escherichia</i> phage vB_EcoM-G3F9	98	99.12	MZ234037.1
EC128	ECOR20	Pond water	Straboviriidae	Tequatrovirus	171814	61340	83	171814	270	8	35.3	ON210139.1	<i>Escherichia</i> phage vB_EcoM_112	94	97.43	KJ668714.2
EC106	ECOR4	Wastewater	None	Felixounavirus	87829	27683	73	87829	128	20	39.1	ON210138.1	<i>Escherichia</i> phage tootiki	91	96.94	MN850647.1
EC150	ECOR36	Cow manure	None	Wifcevirus	68034	601535	2148	68034	94	ND	46.1	ON210137.1	<i>Escherichia</i> phage vB_EcoM-Ro157lw	96	97.53	MH051335.1
EC115	ECOR20	Cow manure	None	Dhillonvirus	44909	299207	1542	44909	63	ND	54.6	ON210136.1	<i>Escherichia</i> phage vB_EcoS_Over9000	98	94.33	OK499985.1

^a ORFs, open reading frames.

^b ND, no tRNAs were identified by RAST annotation.

E. coli phages, a few phages showed limited sequence similarity, thus expanding the knowledge on diversity of *E. coli* phages.

Data availability. The raw sequence reads are available at NCBI SRA database with accession number [SRP394492](https://www.ncbi.nlm.nih.gov/sra/SRP394492) and GenBank accession numbers for the 16 phages are listed in Table 1.

ACKNOWLEDGMENTS

The Danish AgriFish Agency of Ministry of Environment and Food (34009-14-0873) supported S.J.A. and M.G. and Promilleafgiftsfonden supported A.R.V. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. Tenaillon O, Picard B, Denamur E, Skurnik D. 2010. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 8:207–217. <https://doi.org/10.1038/nrmicro2298>.
2. International Committee on Taxonomy of Viruses (ICTV). (2020). <https://ictv.global/msl>.
3. Smith R, O'Hara M, Hobman JL, Millard AD. 2015. Draft genome sequences of 14 *Escherichia coli* phages isolated from cattle slurry. *Gen Announcem (Washington, DC)* 3. <https://doi.org/10.1128/genomeA.01364-15>.
4. Korf IHE, Meier-Kolthoff JP, Adriaenssens EM, Kropinski AM, Nimtz M, Rohde M, van Raaij MJ, Wittmann J. 2019. Still something to discover: novel insights into *Escherichia coli* phage diversity and taxonomy. *Viruses* 11:454. <https://doi.org/10.3390/v11050454>.
5. Olsen NS, Forero-Junco L, Kot W, Hansen LH. 2020. Exploring the remarkable diversity of culturable *Escherichia coli* phages in the Danish wastewater environment. *Viruses* 12:986. <https://doi.org/10.3390/v12090986>.
6. Sørensen PE, Van Den Broeck W, Kiil K, Jasinskyte D, Moodley A, Garmyn A, Ingmer H, Butaye P. 2020. New insights into the biodiversity of coliphages in the intestine of poultry. *Sci Rep* 10:15220–15220. <https://doi.org/10.1038/s41598-020-72177-2>.
7. Maffei E, Shaidullina A, Burkolter M, Heyer Y, Estermann F, Druelle V, Sauer P, Willi L, Michaelis S, Hilbi H, Thaler DS, Harms A, Barr J. 2021. Systematic exploration of *Escherichia coli* phage-host interactions with the BASEL phage collection. *PLoS Biol* 19:e3001424. <https://doi.org/10.1371/journal.pbio.3001424>.
8. Ochman H, Selander RK. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* 157:690–693. <https://doi.org/10.1128/jb.157.2.690-693.1984>.
9. Gencay YE, Gambino M, Prüssing TF, Brøndsted L. 2019. The genera of bacteriophages and their receptors are the major determinants of host range. *Environ Microbiol* 21:2095–2111. <https://doi.org/10.1111/1462-2920.14597>.
10. Sambrook J, Russell DW. 2006. Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc* 2006. <https://doi.org/10.1101/pdb.prot4455>.
11. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards R, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman A, 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* 9:75–90. <https://doi.org/10.1186/1471-2164-9-75>.