



# Complete Genome Sequence of the *Escherichia coli* Phage Ayreon

Marnix Vlot,<sup>a</sup> Franklin L. Nobrega,<sup>b</sup> Che F. A. Wong,<sup>a\*</sup> Yue Liu,<sup>a\*</sup> Stan J. J. Brouns<sup>a,b</sup>

<sup>a</sup>Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, The Netherlands

<sup>b</sup>Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Delft, The Netherlands

**ABSTRACT** We report the whole-genome sequence of a new *Escherichia coli* temperate phage, Ayreon, comprising a linear double-stranded DNA (dsDNA) genome of 44,708 bp.

Phage Ayreon was isolated from pond water samples in Wageningen, The Netherlands (51°58'9.2"N, 5°40'43.1"E) using *Escherichia coli* strain KD471 as the host, a derivative of *E. coli* K-12 derivative strain KD263 (1) lacking clustered regularly interspaced short palindromic repeat (CRISPR)-associated genes *Cas1* and *Cas2*. Infected cultures were given time to allow for lysogeny. Formed lysogens were induced by using both UV exposure and mitomycin C, demonstrating that phage Ayreon is a temperate phage. Transmission electron microscopy revealed icosahedral capsids ( $\pm 57$  nm) and long flexible tails ( $\pm 120$  nm), which are characteristic of *Siphoviridae* (2).

Phage DNA was extracted using the SDS-proteinase K protocol previously described (3). Library preparation and sequencing were performed by BaseClear (The Netherlands) using the Illumina HiSeq 2500 platform. About 1,282,954 short reads were generated with a mean 3,478-fold coverage of the genome. The resulting sequences were *de novo* assembled using the CLC Genomics Workbench version 8.5.1. The contig sequences were corrected with Pilon version 1.11 (4) and linked using the SSPACE Premium scaffold version 2.3 (5). Gapped regions within the scaffolds were partially closed in an automated manner using GapFiller version 1.10 (6). Annotation and identification of open reading frames (ORFs) were performed using the Rapid Annotations using Subsystems Technology (RAST) annotation server (7) followed by manual curation of all predicted proteins against the NCBI protein database using BLASTp (8) and Pfam domain searches (9). tRNAs were predicted with tRNAscan-SE version 1.21 (10), and promoters and terminators were identified using motif searches for TTGACAN(15,18)TATAAT with a maximum of two mismatches and ARNold (11), respectively. The genome packaging strategy was predicted by phylogenetic analysis of the large terminase subunit (12).

Phage Ayreon has a linear double-stranded DNA with a genome size of 44,708 bp and a G+C content of 50.1%, which is very similar to the G+C content of its bacterial host (50.8%). The phage has 59 predicted ORFs, of which 25 could not be assigned to a function. Three predicted promoters and 13 predicted Rho-independent terminators were identified; no tRNAs were identified. Phylogenetic analysis of the large terminase subunit suggested that phage Ayreon uses cohesive end site (*cos*) packaging; *cos* sites are expected to be located within  $\sim 1$  kbp upstream of the small terminase subunit (13). Considering this, the genome of phage Ayreon was opened so that it would begin with the small terminase subunit and end with the expected location of the *cos* site.

The highest degree of similarity was observed with phage cdtI (GenBank accession number AB285204), a cyclomodulin producing prophage (14), with 78% coverage and

Received 2 November 2017 Accepted 21 November 2017 Published 11 January 2018

**Citation** Vlot M, Nobrega FL, Wong CFA, Liu Y, Brouns SJJ. 2018. Complete genome sequence of the *Escherichia coli* phage Ayreon. Genome Announc 6:e01354-17. <https://doi.org/10.1128/genomeA.01354-17>.

**Copyright** © 2018 Vlot 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 Stan J. J. Brouns, [stanbrouns@gmail.com](mailto:stanbrouns@gmail.com).

\* Present address: Che F. A. Wong, Department of Biotechnology and Food Science, Norwegian University of Science and Technology, Trondheim, Norway; Yue Liu, Laboratory of Food Microbiology, Wageningen University, Wageningen, The Netherlands.

96% identity. Phage *cdtI* contains a gene cluster associated with virulence, encoding the CdtA, CdtB, and CdtC subunits of the *cdtI* holotoxin. Whole-genome alignment of phage Ayreon and *cdtI* prophage demonstrates that the *cdtI* gene cluster is absent from phage Ayreon. Global alignment with the *attP* integration site of phage *cdtI* shows pairwise identity of >92%, indicating that phage Ayreon integrates into the gene coding for peptide chain release factor RF-3 in the host genome (14).

**Accession number(s).** The complete genome sequence of Ayreon has been deposited in GenBank under the accession no. [MF807953](https://ncbi.nlm.nih.gov/GenBank/lookup?acc=MF807953).

## ACKNOWLEDGMENTS

This work was supported by The Netherlands Organization for Scientific Research VIDI Grant 864.11.005 (to S.J.J.B.) and Veni Grant 015.Veni-181-092 (to F.L.N.) and by the European Research Council (ERC) Stg grant 638707 (to S.J.J.B.).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

- Shmakov S, Savitskaya E, Semenova E, Logacheva MD, Datsenko KA, Severinov K. 2014. Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res* 42:5907–5916. <https://doi.org/10.1093/nar/gku226>.
- Ackermann HW. 2003. Bacteriophage observations and evolution. *Res Microbiol* 154:245–251. [https://doi.org/10.1016/S0923-2508\(03\)00067-6](https://doi.org/10.1016/S0923-2508(03)00067-6).
- Sambrook J, Russell D. 2001. *Molecular cloning: A laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
- Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol* 13:R56. <https://doi.org/10.1186/gb-2012-13-6-r56>.
- 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. <https://doi.org/10.1186/1471-2164-9-75>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285. <https://doi.org/10.1093/nar/gkv1344>.
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33:W686–W689. <https://doi.org/10.1093/nar/gki366>.
- Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D. 2011. Arnold: a web tool for the prediction of Rho-independent transcription terminators. *RNA Biol* 8:11–13. <https://doi.org/10.4161/rna.8.1.13346>.
- Merrill BD, Ward AT, Grose JH, Hope S. 2016. Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. *BMC Genomics* 17:679. <https://doi.org/10.1186/s12864-016-3018-2>.
- Casjens SR, Gilcrease EB. 2009. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods Mol Biol* 502:91–111. [https://doi.org/10.1007/978-1-60327-565-1\\_7](https://doi.org/10.1007/978-1-60327-565-1_7).
- Asakura M, Hinenoya A, Alam MS, Shima K, Zahid SH, Shi L, Sugimoto N, Ghosh AN, Ramamurthy T, Faruque SM, Nair GB, Yamasaki S. 2007. An inducible lambda-doid prophage encoding cytolethal distending toxin (Cdt-I) and a type III effector protein in enteropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 104:14483–14488. <https://doi.org/10.1073/pnas.0706695104>.