

5-1-2023

Complete genome sequences of evolutionary clade C-III strains of Clostridioides (Clostridium) difficile isolated from the environment in Western Australia

Nirajmohan Shivaperumal

Natasza M. R. Hain-Saunders

Barbara J. Chang

Thomas V. Riley
Edith Cowan University

Daniel R. Knight

Follow this and additional works at: <https://ro.ecu.edu.au/ecuworks2022-2026>



Part of the [Life Sciences Commons](#), and the [Medicine and Health Sciences Commons](#)

[10.1128/mra.00239-23](https://doi.org/10.1128/mra.00239-23)

Shivaperumal, N., Hain-Saunders, N. M. R., Chang, B. J., Riley, T. V., & Knight, D. R. (2023). Complete genome sequences of evolutionary clade C-III strains of Clostridioides (Clostridium) difficile isolated from the environment in Western Australia. *Microbiology Resource Announcements*, 12(5), 1-3. <https://doi.org/10.1128/mra.00239-23>

This Journal Article is posted at Research Online.
<https://ro.ecu.edu.au/ecuworks2022-2026/2734>



Complete Genome Sequences of Evolutionary Clade C-III Strains of *Clostridioides (Clostridium) difficile* Isolated from the Environment in Western Australia

Nirajmohan Shivaperumal,^a Natasza M. R. Hain-Saunders,^b Barbara J. Chang,^a  Thomas V. Riley,^{a,b,c,d}  Daniel R. Knight^{a,c}

^aSchool of Biomedical Sciences, The University of Western Australia, Nedlands, Western Australia, Australia

^bHarry Butler Institute, Murdoch University, Murdoch, Western Australia, Australia

^cDepartment of Microbiology, PathWest Laboratory Medicine WA, Nedlands, Western Australia, Australia

^dSchool of Medical and Health Sciences, Edith Cowan University, Joondalup, Western Australia, Australia

ABSTRACT *Clostridioides (Clostridium) difficile* in the environment is thought to contribute to *C. difficile* infection in community settings. Here, we provide complete genome assemblies for two esculin hydrolysis-negative strains of *C. difficile* that were isolated from soils in Western Australia; the strains produce white colonies on chromogenic media and belong to evolutionarily divergent clade C-III.

Clostridioides (Clostridium) difficile is an anaerobe that can cause life-threatening infectious diarrhea and colitis in humans and nonhuman animals (1). Laboratory detection of *C. difficile* often relies on growth on selective media such as *C. difficile* ChromID agar (bioMérieux, Marcy L'Etoile, France), on which *C. difficile* produces distinctive black colonies (2). However, *C. difficile* strains belonging to PCR ribotype 023 produce colorless or white colonies on ChromID agar and thus may evade laboratory detection (2). Recently, we reported several strains of toxigenic *C. difficile* that were cultured from the environment and that also produced white colonies on chromogenic agar (3, 4), attributable to an inability to hydrolyze esculin. Those strains belonged to clade C-III, a divergent evolutionary lineage that is not well understood (5).

Accurate phylogenetic analyses of *C. difficile* rely on complete reference genomes, which currently exist for all of the major *C. difficile* clades (clades 1 to 5, C-I, and C-II) but not C-III. Here, we provide complete circularized genomes for two esculin hydrolysis-negative strains of *C. difficile* (HGP05 and HGP14) from clade C-III that were isolated from garden soils in Perth, Western Australia, Australia.

C. difficile was originally isolated from soil samples using an anaerobic enrichment culture method, as described previously (4). In preparation for whole-genome sequencing (WGS), pure isolates of *C. difficile* were grown on blood agar for 48 h in an anaerobic chamber (with 80% N₂, 10% CO₂, and 10% H₂). Total genomic DNA for both short-read (Illumina, San Diego, CA, USA) and long-read (Oxford Nanopore Technologies [ONT], Oxford, UK) sequencing was prepared using a QuickGene DNA tissue kit (Kurabo Industries, Osaka, Japan), without shearing or size selection. Short-read WGS was performed using standard Nextera Flex 150-bp paired-end libraries on a NovaSeq 6000 system (Illumina), to an average read depth of 104×. Reads were quality filtered and trimmed using Trim Galore v0.6.7 (<https://github.com/FelixKrueger/TrimGalore>). Long-read sequencing was performed on a MinION Mk1C system (ONT) using an R9 flow cell with a DNA-by-ligation protocol (SQK-LSK109). Base calling was performed using Guppy v4.0.11 integrated within MinKNOW v20.06.17. FilTlong v0.2.0 (<https://github.com/rwrick/Filtlong>) was used to filter low-quality reads (retaining the top 90% of reads), which resulted in 3.07 Gb (HGP05) and 3.09 Gb (HGP14) of sequence data. The Illumina and ONT reads were combined for hybrid genome assembly using Unicycler v0.5.0 (6) in

Editor David Rasko, University of Maryland School of Medicine

Copyright © 2023 Shivaperumal 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 Thomas V. Riley, thomas.riley@uwa.edu.au.

The authors declare no conflict of interest.

Received 22 March 2023

Accepted 12 April 2023

Published 26 April 2023

TABLE 1 Key features of new *C. difficile* genomes

Feature	Data for isolate:	
	HGP05	HGP14
Origin	Soil, Western Australia, 2018	Soil, Western Australia, 2018
PCR ribotype	125	QX597
ST	848 (clade C-III)	632 (clade C-III)
Toxin gene profile ^a	A ⁻ B ⁻ CDT ⁻	A ⁺ B ⁻ CDT ⁻
Chromosome		
GenBank accession no.	CP103977	CP103804
Size (bp)	4,107,441	4,197,547
No. of contigs	1	1
GC content (%)	28.68	28.71
No. of coding sequences	3,693	3,926
No. of tRNAs	91	90
No. of rRNAs	35	35
No. of CRISPRs ^b	7	3
Prophages	ΦCDMH1, ΦMMP01, ΦCDHM19	ΦMMPO3, ΦMMPO1, ΦCDHM19
Extrachromosomal element size (bp) (GenBank accession no.)		
Novel plasmid	5,997 (CP103978) ^c	
Novel phage-like element	63,696 (CP103976) ^d	66,910 (CP103805) ^e , 35,072 (CP103806) ^f
Illumina reads		
Total no. of reads (trimmed)	1,517,442	1,470,894
Avg read length (trimmed) (bp)	150	150
SRA accession no.	SRR16474727	SRR16474718
ONT reads		
Total no. of reads (filtered)	1,846,067	3,280,985
Avg read length (filtered) (bp)	1,800	1,000
Base-called N_{50} (bp)	7,450	4,520
SRA accession no.	SRR23755003	SRR23755002

^a Toxin profile: presence/absence of full-length *tcdA* (A), *tcdB* (pathogenicity locus) (B), and binary toxin *cdtA/B* (binary toxin locus) (CDT).

^b CRISPRs, clusters of regularly interspaced short palindromic repeats.

^c Novel, circularized plasmid showing homology with *C. difficile* pCD6 plasmid (GenBank accession number [NC_005326](#)) (85% sequence identity across 47% of the pCD6 genome).

^d Novel, circularized phage-like element containing a 30.2-kb region showing homology with *C. difficile* phage ΦCDHM19 (GenBank accession number [NC_028996](#)) (75% sequence identity across 18% of the ΦCDHM19 genome).

^e Novel, circularized phage-like element containing a 22.4-kb region showing homology with *C. difficile* phage ΦCDHM19 (GenBank accession number [NC_028996](#)) (75% sequence identity across 15% of the ΦCDHM19 genome). The ΦCDHM19-like regions from HGP05 and HGP14 show significant homology (96% sequence identity across 70% of the sites).

^f Novel, circularized phage-like element containing a 34.0-kb region showing homology with ΦCD111 (GenBank accession number [NC_028905](#)) (89% sequence identity across 66% of the ΦCD111 genome).

conservative assembly mode, with multiple rounds of polishing with Pilon v1.2.4 and Racon v1.4.3 to improve contiguity. Complete circular genomes were confirmed using Bandage v0.8.1 (7) and rotated to *dnaA* using Unicycler. Genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.3 (8). Sequence type (ST) was determined using PubMLST (BIGSdb v1.32.0) (9). Prophages and toxin gene loci were characterized using PHASTER (<https://phaster.ca>) and Abricate v1.0 (<https://github.com/tseemann/abricate>), respectively. Default parameters were used for all software unless otherwise specified.

Information on the assembled chromosomes and extrachromosomal elements is shown in Table 1. We report the complete genome sequences of two esculin hydrolysis-negative strains of *C. difficile*. The white-colony phenotype may be missed in the laboratory; therefore, this data set will aid in future studies of *C. difficile* infection epidemiology and evolution in the community.

Data availability. Complete genome assemblies for HGP05 and HGP14 are available in GenBank under BioProject accession number [PRJNA772357](#) (accession numbers [CP103976](#), [CP103977](#), and [CP103978](#) and accession numbers [CP103804](#), [CP103805](#), and [CP103806](#), respectively). Sequence read data are available in the NCBI Sequence Read Archive (SRA) under BioProject accession number [PRJNA772357](#) (Illumina reads:

SRA accession numbers [SRR16474727](https://www.ncbi.nlm.nih.gov/sra/SRR16474727) [HGP05] and [SRR16474718](https://www.ncbi.nlm.nih.gov/sra/SRR16474718) [HGP14]; Nanopore reads: SRA accession numbers [SRR23755003](https://www.ncbi.nlm.nih.gov/sra/SRR23755003) [HGP05] and [SRR23755002](https://www.ncbi.nlm.nih.gov/sra/SRR23755002) [HGP14]).

ACKNOWLEDGMENTS

N.S. is funded by an Australian Government Research Training Program scholarship. D.R.K. is supported by an Emerging Leader Grant from the Western Australian Department of Health (grant 2022/GR000389). N.M.R.H.-S. is supported by a Research and Innovation Strategic Scholarship from Murdoch University.

REFERENCES

1. Guh AY, Mu Y, Winston LG, Johnston H, Olson D, Farley MM, Wilson LE, Holzbauer SM, Phipps EC, Dumyati GK, Beldavs ZG, Kainer MA, Karlsson M, Gerding DN, McDonald LC. 2020. Trends in U.S. burden of *Clostridioides difficile* infection and outcomes. *N Engl J Med* 382:1320–1330. <https://doi.org/10.1056/NEJMoa1910215>.
2. Connor MC, Fairley DJ, McKenna JP, Marks NJ, McGrath JW. 2016. *Clostridium difficile* ribotype 023 lacks the ability to hydrolyze esculin, leading to false-negative results on chromogenic agar. *J Clin Microbiol* 54:1404–1405. <https://doi.org/10.1128/JCM.00234-16>.
3. Shivaperumal N, Knight DR, Imwattana K, Androga GO, Chang BJ, Riley TV. 2022. Esculin hydrolysis negative and TcdA-only producing strains of *Clostridium (Clostridioides) difficile* from the environment in Western Australia. *J Appl Microbiol* 133:1183–1196. <https://doi.org/10.1111/jam.15500>.
4. Shivaperumal N, Chang BJ, Riley TV. 2020. High prevalence of *Clostridium difficile* in home gardens in Western Australia. *Appl Environ Microbiol* 87:e01572-20. <https://doi.org/10.1128/AEM.01572-20>.
5. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. 2015. Diversity and evolution in the genome of *Clostridium difficile*. *Clin Microbiol Rev* 28:721–741. <https://doi.org/10.1128/CMR.00127-14>.
6. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
7. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31:3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>.
8. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
9. Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, Golubchik T, Harding RM, Jeffery KJM, Jolley KA, Kirton R, Peto TE, Rees G, Stoesser N, Vaughan A, Walker AS, Young BC, Wilcox M, Dingle KE. 2010. Multilocus sequence typing of *Clostridium difficile*. *J Clin Microbiol* 48:770–778. <https://doi.org/10.1128/JCM.01796-09>.