



# Complete Genome Sequences of Four Toxigenic *Clostridium difficile* Clinical Isolates from Patients of the Lower Hudson Valley, New York, USA

Changhong Yin,<sup>a</sup> Donald S. Chen,<sup>b</sup> Jian Zhuge,<sup>c</sup> Donna McKenna,<sup>b</sup> Joan Sagurton,<sup>d</sup> Guiqing Wang,<sup>a,c</sup> Weihua Huang,<sup>a</sup> Nevenka Dimitrova,<sup>e</sup> John T. Fallon<sup>a,c</sup>

<sup>a</sup>Department of Pathology, New York Medical College, Valhalla, New York, USA

<sup>b</sup>Department of Infection Prevention and Control, Westchester Medical Center, Valhalla, New York, USA

<sup>c</sup>Department of Pathology and Clinical Laboratories, Westchester Medical Center, Valhalla, New York, USA

<sup>d</sup>Department of Microbiology, Orange Regional Medical Center, Middletown, New York, USA

<sup>e</sup>Philips Healthcare North America, Cambridge, Massachusetts, USA

**ABSTRACT** Complete genome sequences of four toxigenic *Clostridium difficile* isolates from patients in the lower Hudson Valley, New York, USA, were achieved. These isolates represent four common sequence types (ST1, ST2, ST8, and ST42) belonging to two distinct phylogenetic clades. All isolates have a 4.0- to 4.2-Mb circular chromosome, and one carries a phage.

*Clostridium (Clostridiodes) difficile* is a Gram-positive, spore-forming, anaerobic bacterium that causes antibiotic-associated diarrhea and life-threatening pseudomembranous colitis. The incidence of *C. difficile* infections (CDIs) has dramatically increased in the last decade. It is estimated that approximately 453,000 incident infections with approximately 29,000 deaths occurred in the United States in 2011 (1). The pathogenicity of *C. difficile* is closely linked to two major exotoxins (toxin A and toxin B) encoded by genes *tcdA* and *tcdB*, respectively, in a pathogenicity locus (PaLoc) (2, 3). The virulence of *C. difficile* can be enhanced by producing binary toxin (CDT) (4, 5). Genes encoding the binary toxin, *cdtA* and *cdtB*, together with a regulation gene *cdtR*, are located on the CDT locus (CdtLoc) (6).

Genomic sequence analysis provides insights into the genomic diversity, evolution, and transmission of *C. difficile* (7). However, among 17 *C. difficile* complete genomes publicly available to date in the GenBank database (<https://www.ncbi.nlm.nih.gov/genome/genomes/535>, accessed on 11 December 2017), only 3 (BI1, DH, and FDAARGOS\_267) were from humans in the United States. In this study, we sequenced four *C. difficile* clinical isolates (W0003a, W0022a, W0023a, and R0104a) from patients of the lower Hudson Valley, New York, USA. These isolates were recovered from patients with symptomatic CDIs and represented the most common strains in this region. Total genomic DNA of each isolate was extracted with a Qiagen QIAamp genomic DNA kit. Whole-genome sequencing was performed on both the Pacific Bioscience (PacBio) RS II and Illumina MiSeq platforms. A sequencing library for the MiSeq platform was prepared using the Illumina Nextera XT DNA sample prep kit, and paired-end (2 × 150-bp) sequencing was performed. A library for the PacBio single-molecule real-time (SMRT) sequencing system was processed with the PacBio SMRTbell template preparation kit. Genome sequences were *de novo* assembled by PacBio SMRT analysis and/or the SPAdes assembler (8), polished by aligning and mapping the short reads generated from the MiSeq run, and visualized under the Integrative Genomics Viewer (9).

The genomes of W0003a, W0023a, and R0104a each contain a single circular chromosome of 4,075,361, 4,110,080, and 4,190,038 bp, respectively. W0022a consists

Received 11 December 2017 Accepted 15 December 2017 Published 25 January 2018

**Citation** Yin C, Chen DS, Zhuge J, McKenna D, Sagurton J, Wang G, Huang W, Dimitrova N, Fallon JT. 2018. Complete genome sequences of four toxigenic *Clostridium difficile* clinical isolates from patients of the lower Hudson Valley, New York, USA. Genome Announc 6: e01537-17. <https://doi.org/10.1128/genomeA.01537-17>.

**Copyright** © 2018 Yin 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 Changhong Yin, [changhong\\_yin@nymc.edu](mailto:changhong_yin@nymc.edu).

of a circular chromosome of 4,188,456 bp and a circular phage of 31,888 bp. The sequence type (ST) of each *C. difficile* strain was determined by querying the whole-genome sequence to the public multilocus sequence typing (MLST) database (<http://pubmlst.org/cdifficile>): R0104a (ST1), W0022a (ST2), W0003a (ST8), and W0023a (ST42). The four isolates represent the most common strains recovered in this region. MLST-based phylogeny further designated these isolates into two distinct phylogenetic clades: R0104a as clade 2, and the others as clade 1.

Compared to the reference genome *C. difficile* 630 (GenBank accession number NC\_009089), the genomes of all four toxigenic *C. difficile* isolates contain intact *tcdA* and *tcdB* genes in the PaLoc, of 7,133 and 7,101 bp, respectively. In addition, a complete CdtLoc, including intact binary toxin genes *cdtA* and *cdtB* and their regulatory *cdtR*, was identified in R0104a, whereas an incomplete CdtLoc, containing only the *cdtR* gene, was observed in isolates W0003a, W0022a, and W0023a (5, 10).

**Accession number(s).** Complete genome sequences of the four isolates reported here have been deposited in GenBank under the accession numbers [CP025044](#) to [CP025047](#).

## ACKNOWLEDGMENTS

Funding support was provided in part by the New York State Department of Health, Bureau of Healthcare Associated Infections, grant RFA no. 1203311156, and by Philips Healthcare North America.

## REFERENCES

1. Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JI, Phipps EC, Wilson LE, Winston LG, Cohen JA, Limbago BM, Fridkin SK, Gerding DN, McDonald LC. 2015. Burden of *Clostridium difficile* infection in the United States. *N Engl J Med* 372: 825–834. <https://doi.org/10.1056/NEJMoa1408913>.
2. Loo VG, Poirier L, Miller MA, Oughton M, Libman MD, Michaud S, Bourgault AM, Nguyen T, Frenette C, Kelly M, Vibien A, Brassard P, Fenn S, Dewar K, Hudson TJ, Horn R, René P, Monczak Y, Dasgal A. 2005. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *N Engl J Med* 353:2442–2449. <https://doi.org/10.1056/NEJMoa051639>.
3. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, Frost E, McDonald LC. 2005. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* 366:1079–1084. [https://doi.org/10.1016/S0140-6736\(05\)67420-X](https://doi.org/10.1016/S0140-6736(05)67420-X).
4. Cowardin CA, Buonomo EL, Saleh MM, Wilson MG, Burgess SL, Kuehne SA, Schwan C, Eichhoff AM, Koch-Nolte F, Lyras D, Aktories K, Minton NP, Petri WA, Jr. 2016. The binary toxin CDT enhances *Clostridium difficile* virulence by suppressing protective colonic eosinophilia. *Nat Microbiol* 1:16108. <https://doi.org/10.1038/nmicrobiol.2016.108>.
5. Eckert C, Emirian A, Le Monnier A, Cathala L, De Montclos H, Goret J, Berger P, Petit A, De Chevigny A, Jean-Pierre H, Nebbad B, Camiade S, Meckenstock R, Lalande V, Marchandin H, Barbut F. 2015. Prevalence and pathogenicity of binary toxin-positive *Clostridium difficile* strains that do not produce toxins A and B. *New Microbes New Infect* 3:12–17. <https://doi.org/10.1016/j.nmni.2014.10.003>.
6. Perelle S, Gibert M, Bourlioux P, Corthier G, Popoff MR. 1997. Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infect Immun* 65:1402–1407.
7. 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>.
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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. <https://doi.org/10.1089/cmb.2012.0021>.
9. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29: 24–26. <https://doi.org/10.1038/nbt.1754>.
10. Chen R, Feng Y, Wang X, Yang J, Zhang X, Lü X, Zong Z. 2017. Whole genome sequences of three clade 3 *Clostridium difficile* strains carrying binary toxin genes in China. *Sci Rep* 7:43555. <https://doi.org/10.1038/srep43555>.