



Closed Genome Sequence of *Clostridium botulinum* Strain IBCA10-7060 Type Bh

Nir Dover,^a Jason R. Barash,^a John M. Bell,^a Matthew D. Sylvester,^a  Stephen S. Arnon^a

^aInfectious Diseases Laboratory Branch, Division of Communicable Disease Control, Center for Infectious Diseases, California Department of Public Health, Richmond, California, USA

ABSTRACT *Clostridium botulinum* strain IBCA10-7060 was isolated from a stool specimen from an infant botulism patient and is the only *Clostridium botulinum* strain known that produces botulinum toxin type H. We present here its 4.09-Mbp closed genome sequence.

Clostridium botulinum is a Gram-positive anaerobic, spore-forming bacterium with the ability to produce botulinum neurotoxin, the most potent naturally occurring toxin known (1). Botulinum toxin type H was the first new botulinum toxin type to be recognized in more than 40 years (2, 3), and *Clostridium botulinum* strain IBCA10-7060 is the only *Clostridium* strain known that produces botulinum toxin type H (2–7); this strain also produces botulinum toxin type B (2, 3). Strain IBCA10-7060 was isolated in California from an infant botulism patient, and the stool collection was exempt from institutional review because the testing of specimens from suspected infant botulism patients is required by California law (Health and Safety Code §123704). Previous draft genome sequencing of strain IBCA10-7060 yielded a 4,005,128-bp genome divided into 131 contigs (8).

Strain IBCA10-7060 was subcultured anaerobically from a frozen stock culture on 4% egg yolk agar plates and incubated for 48 h at 35°C. The isolated colonies were inoculated into 20 ml of prereduced Trypticase-peptone-glucose-yeast extract broth, incubated anaerobically for 16 h at 35°C, and harvested by centrifugation at $3,450 \times g$.

DNA was extracted using the Zymo Research (Irvine, CA) Quick-DNA high-molecular-weight (HMW) MagBead kit according to the manufacturer's instructions. For short-read sequencing (300-bp paired-end reads), libraries were prepared using the KAPA HyperPlus kit (Roche, Basel, Switzerland) and sequenced on a MiSeq instrument (Illumina, San Diego, CA). For long-read sequencing, libraries were prepared using the ligation sequencing kit (SQK-LSK109) and sequenced on the MinION Mk1C device on a FLO-MIN106 flow cell for 24 h (Oxford Nanopore, Oxford, UK).

Guppy version 3.6.0 (Oxford Nanopore) was used for base calling to produce a total of 397,436 long reads with an N_{50} value of 31,757 bp. These reads were first filtered using NanoFilt version 2.6 (9) to minimum average read scores of 12, leaving 340,484 reads. The long reads were downsampled four times to produce differing groups of reads with coverage depths of approximately 215-fold. The Illumina short reads were downsampled from 52.8 million reads (26.4 million pairs) to 1.3 million read pairs. All downsampling was done using an in-house python script. Ratatosk version 0.2.3 (10) was run on the filtered long reads to prepolysh them and to improve the quality before assembly. Flye version 2.6 (11), Raven version 0.0.7 (12), Redbean (no version available) (13), and Miniasm version 0.3-r179 (14) were each run on all sets of the samples that had been processed using Ratatosk. These data sets were then processed using Tricycler version 0.3.3 (https://zenodo.org/record/4556428#.YFu_R2hKiUl), which clustered, reconciled, and circularized the contigs, thereby producing a single closed consensus genome sequence of 4,090,796 bp with genome coverage of 405-fold. All programs were run with default

Citation Dover N, Barash JR, Bell JM, Sylvester MD, Arnon SS. 2021. Closed genome sequence of *Clostridium botulinum* strain IBCA10-7060 type Bh. *Microbiol Resour Announc* 10:e00383-21. <https://doi.org/10.1128/MRA.00383-21>.

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Copyright © 2021 Dover 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 Stephen S. Arnon, stephen.arnon@cdph.ca.gov.

Received 6 May 2021

Accepted 28 July 2021

Published 26 August 2021

settings except as noted here: all assemblers were run with 16 threads, Flye and Redbean were given a 4-Mb genome specification, and Flye was given a 1,000 minimum overlap (-m option). Manual inspection indicated a spurious one-base insertion in a homopolymer sequence in *higA*, which was corrected to produce a genome of 4,090,796 bp. The genome sequence was annotated upon submission to NCBI using the Prokaryotic Genome Annotation Pipeline (PGAP) (15).

The closed genome sequence contained 4,090,796 bp (GC content, 28.2%), with 3,714 coding DNA sequences and 9 complete rRNA operons. The genome contained no plasmids. The type H botulinum toxin gene was part of a large 53.7-kb insertion that occurred at a chromosomal location different than that of all other chromosomal or plasmid botulinum toxin gene clusters (3).

Data availability. The closed genome sequence of *Clostridium botulinum* strain IBCA10-7060 has been deposited in GenBank under the accession number [CP069280.1](https://doi.org/10.3390/toxins11120686), the BioProject accession number [PRJNA698972](https://doi.org/10.1093/infdi/jit449), and the BioSample accession number [SAMN17763437](https://doi.org/10.1093/infdi/jit450). The raw sequences were deposited under the SRA accession numbers [SRR14421949](https://doi.org/10.1093/infdi/jit450) and [SRR14421950](https://doi.org/10.1093/infdi/jit450).

ACKNOWLEDGMENT

This work was supported by the Infant Botulism Treatment and Prevention Fund of the California Department of Public Health.

REFERENCES

- Rossetto O, Montecucco C. 2019. Tables of toxicity of botulinum and tetanus neurotoxins. *Toxins (Basel)* 11:686. <https://doi.org/10.3390/toxins11120686>.
- Barash JR, Arnon SS. 2014. A novel strain of *Clostridium botulinum* that produces type B and type H botulinum toxins. *J Infect Dis* 209:183–191. <https://doi.org/10.1093/infdi/jit449>.
- Dover N, Barash JR, Hill KK, Xie G, Arnon SS. 2014. Molecular characterization of a novel botulinum neurotoxin type H gene. *J Infect Dis* 209:192–202. <https://doi.org/10.1093/infdi/jit450>.
- Pellet S, Tepp WH, Lin G, Johnson EA. 2018. Substrate cleavage and duration of action of botulinum neurotoxin type FA (“H, HA”). *Toxicon* 147:38–46. <https://doi.org/10.1016/j.toxicon.2017.12.048>.
- Lam K-H, Sikorra S, Weisemann J, Maatsch H, Perry K, Rummel A, Binz T, Jin R. 2018. Structural and biochemical characterization of the protease domain of the mosaic botulinum neurotoxin type HA. *Pathog Dis* 76:fty044. <https://doi.org/10.1093/femspd/fty044>.
- Fan Y, Barash JR, Lou J, Conrad F, Marks JD, Arnon SS. 2016. Immunological characterization and neutralizing ability of monoclonal antibodies directed against botulinum neurotoxin type H. *J Infect Dis* 213:1606–1614. <https://doi.org/10.1093/infdi/jiv770>.
- Fan Y, Barash JR, Conrad F, Lou J, Tam C, Cheng LW, Arnon SS, Marks JD. 2019. The novel clostridial neurotoxin produced by strain IBCA10-7060 is immunologically equivalent to BoNT/HA. *Toxins (Basel)* 12:9. <https://doi.org/10.3390/toxins12010009>.
- Gonzalez-Escalona N, Thirunavukkarasu N, Singh A, Toro M, Brown EW, Zink D, Rummel A, Sharma SK. 2014. Draft genome sequence of bivalent *Clostridium botulinum* strain IBCA10-7060, encoding botulinum neurotoxin B and a new FA mosaic type. *Genome Announc* 2:e01275-14. <https://doi.org/10.1128/genomeA.01275-14>.
- De Coster W, D’Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- Holley G, Beyter D, Ingimundardottir H, Møller PL, Kristmundsdottir S, Eggertsson HP, Halldorsson BV. 2021. Ratatosk: hybrid error correction of long reads enables accurate variant calling and assembly. *Genome Biol* 22:28. <https://doi.org/10.1186/s13059-020-02244-4>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Vaser R, Šikić M. 2021. Raven: a de novo genome assembler for long reads. *bioRxiv* <https://doi.org/10.1101/2020.08.07.242461>.
- Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* 17:155–158. <https://doi.org/10.1038/s41592-019-0669-3>.
- Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32:2103–2110. <https://doi.org/10.1093/bioinformatics/btw152>.
- Li W, O’Neill KR, Haft DH, DiCuccio M, Chetvermin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res* 49:D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>.