AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequences for Dual-Toxin-Producing *Clostridium botulinum* Strains

Jessica L. Halpin,[a] Janet K. Dykes,[a] Carolina Lúquez[a]

[a]Centers for Disease Control and Prevention, Atlanta, Georgia, USA

ABSTRACT Here, we present draft genome sequences for three *Clostridium botulinum* strains that produce multiple botulinum toxin serotypes. Strains that produce two toxins are rare; however, one of these strains produces subtype B5 and F2 toxins, and two of the strains produce subtype A4 and B5 toxins.

Botulism is a rare but serious neuroparalytic disease that is caused by botulinum neurotoxins (1). These toxins are produced by *Clostridium botulinum* and, rarely, other *Clostridia* species and cause four different types of disease (2), foodborne botulism, infant botulism, wound botulism, and adult intestinal colonization. Rarely, strains may produce more than one toxin type (3–8), and they are denoted as bivalent toxin producers, with the major toxin written as an uppercase letter and the minor toxin as a lowercase letter. Only three combinations of bivalent toxin producers have been identified, as follows: B and F, A and F, and A and B (e.g., Ab, Ba, Bf, and Af).

C. botulinum strain CDC69057 was isolated (9) from a case of infant botulism during routine investigation and determined to be serotype Bf, subtypes B5 and F2. Previous bivalent Bf strains have shared these subtypes (10).

C. botulinum CDC60006 was isolated (9) from a case of infant botulism and determined to be serotype AB, subtypes A4 and B5. *C. botulinum* CDC69043 was isolated from a baked potato during a foodborne botulism investigation and determined to be serotype Ba, subtypes A4 and B5. These two isolates were divided by time and geography and are not believed to be related.

The isolates were retrieved from long-term storage, inoculated into chopped meat glucose starch (CMGS) broth, and grown at 33 to 37°C anaerobically for 24 h. The culture was streaked onto egg yolk agar for isolation. A single colony was inoculated into Trypticase soy glucose yeast (TPGY) broth and grown anaerobically for 16 to 18 h prior to genomic DNA extraction. DNA extraction was performed using a modified MasterPure complete DNA and RNA purification kit (Lucigen, Middleton, WI).

Sequencing was performed on the Ion Torrent Personal Genome Machine (11), with 200-bp libraries created with the Kapa Biosciences kit (12). Raw reads were assessed with FastQC (13) and assembled using SPAdes version 3.1.0 (14), with the following settings enabled: k = 21,33,55,77,99,127; -s; –careful; –iontorrent. Assemblies were assessed using Quast version 4.0 (15). Plasmid SPAdes was run using the same settings plus the –plasmid flag. The traditional 7-gene multilocus sequence type (MLST) (16) was determined by querying assemblies against the Center for Genomic Epidemiology MLST 2.0 site (17). Unicycler version 0.4.4 was run with default settings and the single end flag enabled. The Map Reads to Reference tool was utilized within CLC Genomics Workbench, with default settings, to map the reads to plasmid references.

CDC69057 sequencing resulted in 521,912 reads, with an average size of 170 bp, 23× average coverage, 220 contigs, and an N_{50} value of 34,270 bp. The sequence has

Citation Halpin JL, Dykes JK, Lúquez C. 2019. Draft genome sequences for dual-toxin-producing *Clostridium botulinum* strains. Microbiol Resour Announc 8:e01152-18. https://doi.org/10.1128/MRA.01152-18.

Editor Julia A. Maresca, University of Delaware

Address correspondence to Jessica L. Halpin, JLHalpin@cdc.gov.

Received 21 August 2018
Accepted 19 December 2018
Published 24 January 2019

a GC content of 27.96% and belongs to the 7-gene multilocus sequence type 14 (ST14) (16, 18), which is a very small group. In addition to CDC69057, the PubMLST database contains only 5 isolates with ST14. Four of these isolates are serotype Bf (subtypes B2 and F2) and were isolated during several events in France. One isolate is listed as having an unknown source location and is serotype A(B). Sequencing coverage was not sufficient to confirm the presence of a plasmid, but the reads map with 6.56× average coverage across about 50% of the pCLJ reference from strain CDC657, a dual-toxin-producing strain.

CDC60006 sequencing resulted in 508,444 reads, with an average size of 190 bp, 25× average coverage, 135 contigs, and an N_{50} value of 68,652 bp. CDC69043 sequencing resulted in 794,471 reads, with an average size of 222 bp, 46× average coverage, 93 contigs, and an N_{50} value of 161,283 bp. Both sequences have a GC content of 27.9% and belong to the 7-gene MLST ST7, which is only one allele different from the previously mentioned ST14. ST7 has 23 members on PubMLST from Argentina, France, and the United States and is composed of serotype A in the form of A(B), as well as subtypes A4 and A2. CDC60006 sequencing did not yield conclusive results regarding plasmid presence, possibly due to low coverage. The reads mapped to pCLJ with 10.92× average coverage across about 64% of the reference plasmid from strain CDC657, and these did include toxin genes. Plasmid SPAdes was able to assemble pCLJ2 out of the CDC69043 reads, and Unicycler (19) was able to construct pCLJ, which also included both toxin genes. Plasmids very similar to pCLJ and pCLJ2 reside within CDC69043, and the toxin genes are found within pCLJ.

Data availability. These sequences have been deposited under SRA accession numbers SRS4007694 (CDC60006), SRS4007621 (CDC69057), and SRS4007620 (CDC69043) and GenBank accession numbers QVAF00000000 (CDC60006), QVAG00000000 (CDC69043), and QVAH00000000 (CDC69057).

ACKNOWLEDGMENTS

Support for this project was provided by the CDC Office of Public Health Preparedness and Response.

The findings and conclusions in this report are those of the authors and do not represent the official position of the Centers for Disease Control and Prevention.

We thank the Texas Public Health Laboratory for identifying, isolating, and submitting CDC69057 to our collection.

REFERENCES

1. Cherington M. 1998. Clinical spectrum of botulism. Muscle Nerve 21:701–710. https://doi.org/10.1002/(SICI)1097-4598(199806)21:6<701::AID-MUS1>3.0.CO;2-B.
2. Hatheway CL. 1990. Toxigenic clostridia. Clin Microbiol Rev 3:66–98. https://doi.org/10.1128/CMR.3.1.66.
3. de Jong LIT, Fernández RA, Pareja V, Giaroli G, Guidarelli SR, Dykes JK, Lúquez C. 2015. First report of an infant botulism case due to Clostridium botulinum type Af. J Clin Microbiol 53:740–742. https://doi.org/10.1128/JCM.02894-14.
4. Fernández RA, Ciccarelli AS, Arenas GN, Gimenez DF. 1986. First outbreak of botulism caused by Clostridium botulinum subtype Af. Rev Argent Microbiol 18:29–31. (In Spanish.)
5. Giménez DF, Ciccarelli AS. 1978. New strains of Clostridium botulinum subtype Af. Zentralbl Bakteriol Orig A 240:215–220.
6. Smith TJ, Hill KK, Foley BT, Detter JC, Munk AC, Bruce DC, Doggett NA, Smith LA, Marks JD, Xie G, Brettin TS. 2007. Analysis of the neurotoxin complex genes in Clostridium botulinum A1-A4 and B1 strains: BoNT/A3,/Ba4 and/B1 clusters are located within plasmids. PLoS One 2:e1271. https://doi.org/10.1371/journal.pone.0001271.
7. Giménez DF. 1984. Clostridium botulinum subtype Ba. Zentralbl Bakteriol Mikrobiol Hyg A 257:68–72.
8. Barash JR, Arnon SS. 2004. Dual toxin-producing strain of Clostridium botulinum type Bf isolated from a California patient with infant botulism. J Clin Microbiol 42:1713–1715. https://doi.org/10.1128/JCM.42.4.1713-1715.2004.
9. Centers for Disease Control and Prevention. 1998. Botulism in the United States (1899–1996). Handbook for epidemiologists, clinicians, and laboratory workers. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/botulism/pdf/bot-manual.pdf.
10. Raphael BH, Choudoir MJ, Luquez C, Fernandez R, Maslanka SE. 2010. Sequence diversity of genes encoding botulinum neurotoxin type F. Appl Environ Microbiol 76:4805–4812. https://doi.org/10.1128/AEM.03109-09.
11. Thermo Fisher Scientific. 2017. Ion Personal Genome Machine (PGM) system reference guide, vol MAN0009783. Rev. A.0. Life Technologies Corporation, Carlsbad, CA. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0009783_IonPGM_RefGuide.pdf.
12. Kapa Biosystems. 2016. Kapa library preparation kit. Ion Torrent platforms, v2.16. Kapa Biosystems, Wilmington, MA.
13. Andrews S. 2014. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
14. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 20:714–737. https://doi.org/10.1089/cmb.2013.0084.
15. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
16. Jacobson MJ, Lin G, Whittam TS, Johnson EA. 2008. Phylogenetic analysis

of Clostridium botulinum type A by multi-locus sequence typing. Microbiology 154:2408–2415. https://doi.org/10.1099/mic.0.2008/016915-0.
17. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Ponten T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361. https://doi.org/10.1128/JCM.06094-11.
18. Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. https://doi.org/10.1186/1471-2105-11-595.
19. 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.