AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Closed Genome Sequences of Two *Clostridium botulinum* Strains Obtained by Nanopore Sequencing

Narjol Gonzalez-Escalona,[a] Julie Haendiges,[a,b] Jesse D. Miller,[b] Shashi K. Sharma[a]

[a]Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, Maryland, USA

[b]NSF International, Applied Research Center, Ann Arbor, Michigan, USA

ABSTRACT Here we report the genome sequences of two toxin-producing *Clostridium botulinum* strains, one environmental sample (83F) and one clinical sample (CDC51232). The genomes were closed by a combination of long-read and short-read sequencing. The strains belong to *C. botulinum* sequence type 4 (ST4) and ST7, respectively.

Clostridium botulinum is a Gram-positive, spore-forming anaerobic bacterium that produces botulinum neurotoxin (BoNT) (1). Ingestion of the potent BoNT causes a serious paralytic illness known as botulism in humans and is a critical concern for food safety. The neurotoxins produced by these organisms are serologically different, and seven serotypes have been described, designated by the letters A through G (2). Four of the seven serotypes, namely, A, B, E, and F, have been linked with human botulism, with most cases due to serotypes A and B (3).

The genomes of two bivalent toxin-producing *C. botulinum* strains (strains that carry two botulinum toxins) were sequenced to be prepared for botulism outbreaks. The strains were grown, and the DNA was extracted as reported previously (4). The long reads for each strain were generated with MinION sequencing (Nanopore, Oxford, UK). The sequencing libraries were prepared using the rapid sequencing kit RAD004 and run in a FLO-MIN106 (R9.4.1) flow cell, according to the manufacturer's instructions, for 48 h at 230 to 290× average coverage. The sequencing library contained DNA fragmented randomly by a transposase present in the fragmentation mix of the RAD004 kit, rendering fragments of >30 kb. The short-read whole-genome sequence (WGS) for each strain was generated using the Illumina MiSeq sequencing platform with the MiSeq v3 kit using 2 × 250-bp paired-end chemistry (Illumina, San Diego, CA) according to the manufacturer's instructions at 160 to 180× coverage. The libraries were constructed with 100 ng of genomic DNA using the Nextera DNA flex kit (Illumina) according to the manufacturer's instructions. The genome sequences for each strain were obtained by *de novo* assembly, using nanopore data and default settings within the Canu program v1.6 (5). A second assembly was generated using a SPAdes (6) hybrid assembly (with default settings) using both Nanopore and MiSeq data generated for each strain. The resultant assemblies from Canu were error corrected using the Pilon tool (7) and the MiSeq data. The final assembly (FA) was generated by comparing the SPAdes hybrid and Canu-polished assemblies using Mauve (8) and filling in the missing regions in the SPAdes assembly with the Canu-polished assembly. The FA sequences were annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP, https://www.ncbi.nlm.nih.gov/genome/annotation_prok).

In silico multilocus sequence typing (MLST) analyses (https://pubmlst.org/bigsdb?db=pubmlst_cbotulinum_seqdef&page=sequenceQuery) showed that CDC51232 belonged to sequence type 7 (ST7) and 83F belonged to ST4. Whole-genome single-nucleotide polymorphism (SNP) analysis, performed as described previously (4),

Received 31 July 2018 **Accepted** 3 August 2018 **Published** 6 September 2018

Citation Gonzalez-Escalona N, Haendiges J, Miller JD, Sharma SK. 2018. Closed genome sequences of two *Clostridium botulinum* strains obtained by Nanopore sequencing. Microbiol Resour Announc 7:e01075-18. https://doi.org/10.1128/MRA.01075-18.

Editor J. Cameron Thrash, Louisiana State University

Address correspondence to Narjol Gonzalez-Escalona, narjol.gonzalez-escalona@fda.hhs.gov.

TABLE 1 Metadata for the two *C. botulinum* strains reported in this study[a]

CFSAN no.	Isolate name	GenBank accession no. (size [bp])		Sequence Read Archive no.	Source	Serotype	Sequence type
		Chromosome	Plasmids				
CFSAN034200	CDC51232	CP031095 (270,024)	CP031096 (9,953)	SRR7530166	Clinical	AB	7
			CP031097 (3,922,194)	SRR7530167			
CFSAN034202	83F	CP031098 (3,954,901)	CP031100 (57,676)	SRR7532471	Environmental	AB	4
			CP031099 (5,926)	SRR7532470			

[a]The GC content for each strain was 28.2%.

showed that these genomes belonged to two different lineages, with CDC51232 and 83F belonging to lineages 2 and 4, respectively, and contain mostly bivalent strains, as inferred from our previous study (4). Analysis of the resulting sequences showed the presence of two plasmids in each sequenced strain, although the sizes and sequences of these two plasmids differed greatly between each other (Table 1). Furthermore, although these two isolates were also bivalent *C. botulinum* strains, the location of the BoNT clusters differed between them. In strain CDC51232 the BoNT clusters (BoNTB and BoNTA4) were located in the larger plasmid, whereas in 83F, the BoNT clusters (BoNTB and BoNTA1) were located in the chromosome (Table 1).

Data availability. The genome sequences of the two *C. botulinum* strains are listed in Table 1.

ACKNOWLEDGMENTS

This study was supported by funding from the MCMi Challenge grants program proposal no. 2018-646 and FDA Foods Program intramural funds. J.H. and J.D.M. were supported by the NSF International Applied Research Center.

REFERENCES

1. Gill DM. 1982. Bacterial toxins: a table of lethal amounts. Microbiol Rev 46:86–94.
2. Shapiro RL, Hatheway C, Swerdlow DL. 1998. Botulism in the United States: a clinical and epidemiologic review. Ann Intern Med 129:221–228.
3. Kubota T, Yonekura N, Hariya Y, Isogai E, Isogai H, Amano K-I, Fujii N. 1998. Gene arrangement in the upstream region of *Clostridium botulinum* type E and *Clostridium butyricum* BL6340 progenitor toxin genes is different from that of other types. FEMS Microbiol Lett 158:215–221. https://doi.org/10.1111/j.1574-6968.1998.tb12823.x.
4. Gonzalez-Escalona N, Timme R, Raphael BH, Zink D, Sharma SK. 2014. Whole-genome single-nucleotide-polymorphism analysis for discrimination of *Clostridium botulinum* group I strains. Appl Environ Microbiol 80:2125–2132. https://doi.org/10.1128/AEM.03934-13.
5. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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.
7. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
8. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.