# Draft Genome Sequence of a *Clostridium botulinum* Isolate from Thailand Harboring the Subtype *bont*/B8 Gene

Jessica L. Halpin,[a] Piyada Wangroongsarb,[b] Chutima Jittaprasartsin,[b] Janet K. Dykes,[a] Carolina Lúquez[a]

aNational Botulism and Enteric Toxins Team, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Atlanta, Georgia, USA
bNational Institute of Health, Department of Medical Sciences, Ministry of Public Heath, Nonthaburi, Thailand

**ABSTRACT** In 2010, a *Clostridium botulinum* type B isolate was recovered from fermented soybeans during a foodborne botulism investigation. Molecular investigation of the botulinum neurotoxin (*bont*) gene operon determined that the sequence was a new subtype, denoted B8. Here, we describe the draft whole-genome sequence of the organism.

During a foodborne outbreak investigation in Thailand in 2010, *Clostridium botulinum* type B was isolated from fermented soybeans. In 2014, an extensive description of the outbreak, the botulinum toxin (*bont*) gene, and its amino acid sequence was published (1). The authors determined that the toxin gene produced a unique amino acid sequence denoted *bont*/B8 (2, 3). Until now, the genome sequence of the full organism has not been published in a draft or complete form. The whole-genome sequence is important, as the isolate serves as a type strain to represent subtype *bont*/B8, and here we present the draft assembly and the short read data.

The isolate was inoculated into chopped meat glucose starch broth and grown at 35°C anaerobically for 24 h. Culture was streaked onto egg yolk agar for isolation of single colonies, and a single colony was inoculated into Trypticase soy glucose yeast broth and grown at 35°C 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 S5 sequencer (Thermo Fisher, Waltham, MA) with 400-bp libraries created with the Kapa Biosciences kit (Roche, Wilmington, MA). Raw reads (284-bp average) were assessed using FastQC v. 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and assembled using SPAdes v. 3.1.0 (4) with kmer values set to 21, 33, 55, 77, 99, and 127 and with the iontorrent, careful, and single-ended flags activated. The resulting assembly was assessed using QUAST v. 4.0 (5).

The draft assembly consisted of a length of 4,162,303 bp represented by 35.7× average coverage, 569 contigs, and an $N_{50}$ value of 13,659 bp. The sequence has a GC content of 27.96% and represents a new 7-loci multilocus sequence type (6, 7) designated ST-108. The assembly was annotated with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v. 4.6 upon upload to NCBI and resulted in the identification of 4,443 genes, 4,311 coding sequences (CDS), 132 RNA genes, and 1,139 pseudogenes.

A complete plasmid was not assembled, but there is evidence of a plasmid. Mapping reads to plasmid references showed that 4,873 reads mapped to reference plasmid pBOT3502 (GenBank accession number NC_009496). This represented 55% of the total length of 16,344 bp and was covered with good depth (average of 82.29×). These reads contain the toxin gene *bont*/B8.

**Data availability.** The reads were deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRP128500. A draft assembly was also deposited at GenBank under the accession number QVOC00000000.

## ACKNOWLEDGMENTS

## REFERENCES

1. Wangroongsarb P, Kohda T, Jittaprasartsin C, Suthivarakom K, Kamthalang T, Umeda K, Sawanpanyalert P, Kozaki S, Ikuta K. 2014. Molecular characterization of Clostridium botulinum isolates from foodborne outbreaks in Thailand, 2010. PLoS One 9:e77792. https://doi.org/10.1371/journal.pone.0077792.
2. Smith TJ, Hill KK, Raphael BH. 2015. Historical and current perspectives on Clostridium botulinum diversity. Res Microbiol 166:290–302. https://doi.org/10.1016/j.resmic.2014.09.007.
3. Peck MW, Smith TJ, Anniballi F, Austin JW, Bano L, Bradshaw M, Cuervo P, Cheng LW, Derman Y, Dorner BG, Fisher A, Hill KK, Kalb SR, Korkeala H, Lindström M, Lista F, Lúquez C, Mazuet C, Pirazzini M, Popoff MR, Rossetto O, Rummel A, Sesardic D, Singh BR, Stringer SC. 2017. Historical perspectives and guidelines for botulinum neurotoxin subtype nomenclature. Toxins 9:38. https://doi.org/10.3390/toxins9010038.
4. 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.
5. 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.
6. 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.
7. 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.