genomeA
Journals.ASM.org

First Complete Genome Sequence of *Clostridium sporogenes* DSM 795^T, a Nontoxigenic Surrogate for *Clostridium botulinum*, Determined Using PacBio Single-Molecule Real-Time Technology

Kazuma Nakano, Yasunobu Terabayashi, Akino Shiroma, Makiko Shimoji, Hinako Tamotsu, Noriko Ashimine, Shun Ohki, Misuzu Shinzato, Kuniko Teruya, Kazuhito Satou, Takashi Hirano

Okinawa Institute of Advanced Sciences, Uruma, Okinawa, Japan

The first complete genome sequence of *Clostridium sporogenes* DSM 795^T, a nontoxigenic surrogate for *Clostridium botulinum*, was determined in a single contig using the PacBio single-molecule real-time technology. The genome (4,142,990 bp; G+C content, 27.98%) included 86 sets of >1,000-bp identical sequence pairs and 380 tandem repeats.

Received 18 June 2015 **Accepted** 22 June 2015 **Published** 30 July 2015

Citation Nakano K, Terabayashi Y, Shiroma A, Shimoji M, Tamotsu H, Ashimine N, Ohki S, Shinzato M, Teruya K, Satou K, Hirano T. 2015. First complete genome sequence of *Clostridium sporogenes* DSM 795^T, a nontoxigenic surrogate for *Clostridium botulinum*, determined using PacBio single-molecule real-time technology. Genome Announc 3(4): e00832-15. doi:10.1128/genomeA.00832-15.

Address correspondence to Kazuma Nakano, nakano@oias.or.jp.

Clostridium sporogenes is an anaerobic spore-forming bacterium that causes food spoilage (1, 2). *C. sporogenes* is widely used as a nontoxigenic surrogate for *Clostridium botulinum* in the validation of food sterilization because of its physiological and phylogenetic similarity to *C. botulinum* and nontoxigenicity (2–6).

A draft sequence of *C. sporogenes* DSM 795^T has been determined using 454, Illumina, and Sanger technologies in 16 contigs (GenBank accession number JFBQ00000000) (total 4,106,665 bp; average G+C content, 27.8%) (A. Poehlein, R. Karin, S. M. Koenig, R. Daniel, and P. Duerre, submitted for publication) (7, 8). These contigs are disconnected at tandem repeat or low G+C regions. Here, we report the first complete genome sequence of *C. sporogenes* DSM 795^T determined using the PacBio single-molecule real-time (SMRT) technology (9).

The genomic DNA of *C. sporogenes* DSM 795^T, originally isolated from soil in 1908, was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) (10). It was purified using a PowerClean DNA cleanup kit (MoBio, Carlsbad, CA), followed by a 20-kb library construction for P5-C3 chemistry. After >7-kb size selection using BluePippin (Sage Science, Beverly, MA), 8 SMRT cells from the libraries were sequenced using the PacBio RS II platform (Pacific Biosciences, Menlo Park, CA) with 180-min movies. *De novo* assembly was performed using the hierarchical genome assembly process 2 (HGAP2) workflow (11). A single circular contig representing a chromosome was obtained (4,142,990 bp; average G+C content, 27.98%).

The complete genome sequence of *C. sporogenes* DSM 795^T included 86 sets of >1,000-bp identical sequence pairs (4,911-bp maximum) and 380 tandem repeats (369 bp × 8.5 copies maximum). Tandem repeats were identified using Tandem Repeats Finder (12). Recently, a sequence of *C. sporogenes* NCIMB 10696^T, which originated from the same strain (McClung 2004^T) as *C. sporogenes* DSM 795^T, has been determined using 454, Illumina, and Sanger technologies (CP009225) (http://www.straininfo.net/strains/7982) (4,141,984 bp; average G+C content, 28.00%) (13). We found three marked differences between the sequences of DSM 795^T and NCIMB 10696^T. First, in a 39-bp tandem region, DSM 795^T carried 25.5 copies (1,156,066 to 1,157,028), whereas 10696^T carried 20.5 copies (1,156,066 to 1,156,839). Second, in a 312-bp tandem region, DSM 795^T carried 5.9 copies (3,502,125 to 3,503,970), whereas 10696^T carried 4.9 copies (3,501,430 to 3,502,963). Third, DSM 795^T had a 501-bp extra region (2,040,199 to 2,040,699) that could be inserted in 10696^T (between 2,040,006 and 2,040,007). On DSM 795^T sequencing, the PacBio RS II platform produced extra-long reads with an average of 3,959 bp and a maximum of 35,904 bp, and large numbers of reads completely covered those regions: 290 reads for the first, 191 reads for the second, and 359 reads for the third. This result suggests that the number of tandem repeats is underestimated in the 10696^T sequence. The SMRT technology provides power for genome sequencing with multikilobase extra-long reads and unbiased G+C coverage (11, 14, 15) for assessing structural variations such as variable number tandem repeat.

Nucleotide sequence accession number. The complete genome sequence of *C. sporogenes* DSM 795^T was deposited in DDBJ/ENA/GenBank under the accession number CP011663.

ACKNOWLEDGMENT

This work was supported by the Okinawa Prefectural Government.

REFERENCES

1. **Mcclure PJ**. 2006. Spore-forming bacteria, p 579–623. *In* Blackburn (ed), Food spoilage microorganisms, vol 21. Woodhead Publishing, Sawston, United Kingdom.
2. **Brunt J, Plowman J, Gaskin DJ, Itchner M, Carter AT, Peck MW**. 2014. Functional characterisation of germinant receptors in *Clostridium botulinum* and *Clostridium sporogenes* presents novel insights into spore germination systems. PLoS Pathog **10:**e1004382. http://dx.doi.org/10.1371/journal.ppat.1004382.
3. **Bradbury M, Greenfield P, Midgley D, Li D, Tran-Dinh N, Vriesekoop F, Brown JL**. 2012. Draft genome sequence of *Clostridium sporogenes* PA 3679, the common nontoxigenic surrogate for proteolytic *Clostridium*

botulinum. J Bacteriol **194:**1631–1632. http://dx.doi.org/10.1128/JB.06765-11.

4. **Taylor RH, Dunn ML, Ogden LV, Jefferies LK, Eggett DL, Steele FM**. 2013. Conditions associated with *Clostridium sporogenes* growth as a surrogate for *Clostridium botulinum* in nonthermally processed canned butter. J Dairy Sci **96:**2754–2764. http://dx.doi.org/10.3168/jds.2012-6209.
5. **Brown JL, Tran-Dinh N, Chapman B**. 2012. *Clostridium sporogenes* PA 3679 and its uses in the derivation of thermal processing schedules for low-acid shelf-stable foods and as a research model for proteolytic *Clostridium botulinum*. J Food Prot **75:**779–792. http://dx.doi.org/10.4315/0362-028X.JFP-11-391.
6. **Kalia VC, Mukherjee T, Bhushan A, Joshi J, Shankar P, Huma N**. 2011. Analysis of the unexplored features of *rrs* (16S rDNA) of the genus *Clostridium*. BMC Genomics **12:**18. http://dx.doi.org/10.1186/1471-2164-12-18.
7. **Skerman VBD, McGowan V, Sneath PHA**. 1980. Approved lists of bacterial names. Int J Syst Bacteriol **30:**225–420. http://dx.doi.org/10.1099/00207713-30-1-225.
8. **Metchnikoff E**. 1908. Etudes sur la flore intestinale. Ann Inst Pasteur **22:**922–955.
9. **Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S**. 2009. Real-time DNA sequencing from single polymerase molecules. Science **323:**133–138. http://dx.doi.org/10.1126/science.1162986.
10. **Barash JR, Hsia JK, Arnon SS**. 2010. Presence of soil-dwelling clostridia in commercial powdered infant formulas. J Pediatr **156:**402–408. http://dx.doi.org/10.1016/j.jpeds.2009.09.072.
11. **Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J**. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10:**563–569. http://dx.doi.org/10.1038/nmeth.2474.
12. **Benson G**. 1999. Tandem Repeats Finder: a program to analyze DNA sequences. Nucleic Acids Res **27:**573–580. http://dx.doi.org/10.1093/nar/27.2.573.
13. **Zhang Y, Grosse-Honebrink A, Minton NP**. 2015. A universal *mariner* transposon system for forward genetic studies in the genus *Clostridium*. PLoS One **10:**e0122411. http://dx.doi.org/10.1371/journal.pone.0122411.
14. **Shin SC, Ahn do H, Kim SJ, Lee H, Oh T-J, Lee JE, Park H**. 2013. Advantages of single-molecule real-time sequencing in high-GC Content Genomes. PLoS One **8:**e68824. http://dx.doi.org/10.1371/ journal-.pone.0068824.
15. **Satou K, Shiroma A, Teruya K, Shimoji M, Nakano K, Juan A, Tamotsu H, Terabayashi Y, Aoyama M, Teruya M, Suzuki R, Matsuda M, Sekine A, Kinjo N, Kinjo F, Yamaoka Y, Hirano T**. 2014. Complete genome sequences of eight *Helicobacter pylori* strains with different virulence factor genotypes and methylation profiles, isolated from patients with diverse gastrointestinal diseases on Okinawa Island, Japan, determined using PacBio single-molecule real-time technology. Genome Announc **2**(2)**:** e00286-14. http://dx.doi.org/10.1128/genomeA.00286-14.