

Complete Genome Sequence of the Nonpathogenic Soil-Dwelling Bacterium *Clostridium sporogenes* Strain NCIMB 10696

Aleksandra M. Kubiak,^a Anja Poehlein,^b Patrick Budd,^a Sarah A. Kuehne,^a Klaus Winzer,^a Jan Theys,^c Philip Lambin,^c Rolf Daniel,^b Nigel P. Minton^a

The Clostridia Research Group, BBSRC/EPSC Synthetic Biology Research Centre, School of Life Sciences, Centre for Biomolecular Sciences, University of Nottingham, Nottingham, United Kingdom^a; Georg-August-University Goettingen Institute of Microbiology and Genetics, Department of Genomic and Applied Microbiology, Goettingen Genomics Laboratory, Göttingen, Germany^b; Maastr Lab, Maastricht Radiation Oncology, University of Maastricht, Maastricht, The Netherlands^c

***Clostridium sporogenes* is a harmless spore-forming anaerobe that is widely distributed in soil/water and in the intestines of humans and animals. It is extensively used as a safe model to test the suitability of new preservative methods by the food industry and has potential to deliver therapeutic agents to tumors.**

Received 12 July 2015 Accepted 14 July 2015 Published 20 August 2015

Citation Kubiak AM, Poehlein A, Budd P, Kuehne SA, Winzer K, Theys J, Lambin P, Daniel R, Minton NP. 2015. Complete genome sequence of the nonpathogenic soil-dwelling bacterium *Clostridium sporogenes* strain NCIMB 10696. *Genome Announc* 3(4):e00942-15. doi:10.1128/genomeA.00942-15.

Copyright © 2015 Kubiak et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Nigel P. Minton, nigel.minton@nottingham.ac.uk.

Clostridium sporogenes NCIMB 10696 has the potential to be a tumor delivery vehicle for therapeutic agents (1, 2). It is classified by the UK Advisory Committee on Dangerous Pathogens as a harmless group I organism and by the American Type Culture Collection (ATCC 3584) as a harmless biosafety level 1 organism. Its categorization as safe is further demonstrated by the routine use of its spores by the food industry to assess the effectiveness of food preservation methods (3). While closely related to *Clostridium botulinum*, *C. sporogenes* does not produce botulinum toxin and differs markedly with regard to sporulation kinetics (4).

Genomic DNA was prepared by phenol-chloroform extraction (5) and sequenced using the 454 Titanium FLX (Roche Diagnostics) with 3 separate paired-end read sequencing libraries. Genomic DNA was also Illumina resequenced by GATC Biotech (Germany). The sequencing data (1,983,599 reads) were assembled using the GS assembler (Newbler 2.3) into 54 contigs (>500 bp) with a total size of 4,080,182 bp providing 74× coverage. The average contig length was 75,558 bp and the average scaffold size of 682,834 bp, with the largest scaffold being 3,498,261 bp. A total of 39 contig gaps within scaffolds were closed using standard PCR procedures, leading to a draft genome of 4,087,164 bp composed of 5 large scaffolds. Scaffold order was predicted using the move contigs tool of the Mauve Genome Alignment Software (6) and via Gene Ortholog Neighborhoods based on bidirectional best hits implemented at the IMG-ER (Integrated Microbial Genomes/Expert Review) system (7). Alignments were performed using the genomes of *C. sporogenes* ATCC 15579 (GenBank accession number ABKW00000000), *C. botulinum* ATCC 3502, and *C. botulinum* BoNT/B1 Okra (accession number CP000939) as references. A repeat of the assembly was carried out using the Newbler and Mira v3.4 packages and additional Sanger dideoxy sequencing. The final genome sequence length is 4,141,984 bp with an average coverage of 115×.

Gene prediction was performed using the software tool Prodigal (8). Genes encoding rRNA and tRNA were identified using

RNAmmer (9) and tRNAscan (10), respectively. The IMG/ER (Integrated Microbial Genomes/Expert Review) system (7) was used for automatic annotation. Subsequent manual curation used the Swiss-Prot, TrEMBL, and InterPro databases (11). The number of annotated features are coding sequence (CDS), 3,732; protein-encoding genes with function prediction, 2,956; putative genes coding for hypothetical proteins, 776; pseudogenes, 4; rRNA clusters, 9 (27 genes); and tRNA, 80.

The G+C content was 27.98%, and 82% of the genome was encoding. BLAST comparison revealed a 97 to 100% match to other *C. sporogenes* species and 91 to 99% to proteolytic *C. botulinum* strains (12, 13). The 16S rRNA gene sequences of *C. botulinum* and *C. sporogenes* strains share 99 to 100% nucleotide similarity. The assembly and annotation did not reveal the presence of any induced prophages or plasmids. BLAST and PCR analysis of the major toxins produced by *C. botulinum* (groups I through IV) demonstrated that no toxin-encoding genes or remnants are present, highlighting the nonpathogenic nature of *C. sporogenes* NCIMB 10696 and supporting its use in clostridial-directed enzyme prodrug therapy (CDEPT) (14, 15).

Nucleotide sequence accession number. The genome sequence is deposited at GenBank under the accession number CP009225.

ACKNOWLEDGMENTS

This work was supported by the University of Nottingham and the Netherlands Organisation for Health Research and Development, ZonMw.

We thank Martin J. Blythe and Sunir Malla, Deep Seq, Centre for Genetics and Genomics, Queen's Medical Center, University of Nottingham, Nottingham, United Kingdom.

REFERENCES

- Minton NP. 2003. Clostridia in cancer therapy. *Nat Rev Microbiol* 1:237–242. <http://dx.doi.org/10.1038/nrmicro777>.
- Theys J, Pennington O, Dubois L, Anlezark G, Vaughan T, Mengesha A, Landuyt W, Anné J, Burke PJ, Dürre P, Wouters BG, Minton NP,

- Lambin P. 2006. Repeated cycles of *Clostridium*-directed enzyme prodrug therapy result in sustained antitumour effects *in vivo*. *Br J Cancer* 95: 1212–1219. <http://dx.doi.org/10.1038/sj.bjc.6603367>.
3. Brown KL. 2000. Control of bacterial spores. *Br Med Bull* 56:158–171. <http://dx.doi.org/10.1258/0007142001902860>.
 4. Bull MK, Olivier SA, van Diepenbeek RJ, Kormelink F, Chapman B. 2009. Synergistic inactivation of spores of proteolytic *Clostridium botulinum* strains by high pressure and heat is strain and product dependent. *Appl Environ Microbiol* 75:434–445. <http://dx.doi.org/10.1128/AEM.01426-08>.
 5. Sambrook J, Russell DR. 2001. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 6. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <http://dx.doi.org/10.1101/gr.2289704>.
 7. Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I, Tringe S, Huntemann M, Billis K, Varghese N, Tennessen K, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC. 2014. IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* 42:D568–D573. <http://dx.doi.org/10.1093/nar/gkt919>.
 8. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
 9. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
 10. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
 11. Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848. <http://dx.doi.org/10.1093/bioinformatics/17.9.847>.
 12. 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 nontoxicogenic surrogate for proteolytic *Clostridium botulinum*. *J Bacteriol* 194:1631–1632. <http://dx.doi.org/10.1128/JB.06765-11>.
 13. Sebahia M, Peck MW, Minton NP, Thomson NR, Holden MT, Mitchell WJ, Carter AT, Bentley SD, Mason DR, Crossman L, Paul CJ, Ivens A, Wells-Bennik MH, Davis IJ, Cerdeño-Tárraga AM, Churcher C, Quail MA, Chillingworth T, Feltwell T, Fraser A, Goodhead I, Hance Z, Jagels K, Larke N, Maddison M, Moule S, Mungall K, Norbertczak H, Rabinowitsch E, Sanders M, Simmonds M, White B, Whithead S, Parkhill J. 2007. Genome sequence of a proteolytic (group I) *Clostridium botulinum* strain hall A and comparative analysis of the clostridial genomes. *Genome Res* 17:1082–1092. <http://dx.doi.org/10.1101/gr.6282807>.
 14. Kubiak AM, Minton NP. 2015. The potential of clostridial spores as therapeutic delivery vehicles in tumour therapy. *Res Microbiol* 166: 244–254. <http://dx.doi.org/10.1016/j.resmic.2014.12.006>.
 15. Heap JT, Theys J, Ehsaan M, Kubiak AM, Dubois L, Paesmans K, Mellaert LV, Knox R, Kuehne SA, Lambin P, Minton NP. 2014. Spores of *Clostridium* engineered for clinical efficacy and safety cause regression and cure of tumors *in vivo*. *Oncotarget* 5:1761–1769.