Draft Genome Sequence of the Cellulolytic Strain *Clostridium* sp. Bc-iso-3 Isolated from an Industrial-Scale Anaerobic Digester

Li Sun, Anna Schnürer

Department of Microbiology, Swedish University of Agricultural Science, Uppsala BioCenter, Uppsala, Sweden

Clostridium sp. Bc-iso-3 is a cellulolytic strain isolated from a Swedish industrial-scale biogas digester. Here, we present the draft genome sequence of this strain, which consists of four contigs with a total length of 4,327,139 bp and an average coverage of 312.97×.

Address correspondence to Li Sun, li.sun@slu.se, or Anna Schnürer, anna.schnurer@slu.se.

Cellulosic materials represent the most abundant biopolymer on earth (1), and microbial cellulose utilization is a key step in the global carbon cycle (2). Consequently, cellulases produced by microorganisms are not only of importance in nature but are also important for various applications in industrial processes (3). Anaerobic digestion represents one such interesting application, and during this process organic materials can be converted to renewable energy in the form of biogas. An array of different microbial steps is required for this process, and the first step, the hydrolysis, represents a key step during the degradation of plant-based materials (4). With the aim to further understand the hydrolysis of lignocellulose in biogas digester isolations were performed. A cellulolytic strain, *Clostridium* sp. Bc-iso-3, was retrieved from an industrial anaerobic digester (CSTR) in Sweden. Based on a full-length 16S rRNA gene analysis, its closest relative is *Clostridium thermocellum* DSM 2360 with 96% sequence identity. At the time of sampling, this digester was operated at 38°C using source-separated municipal organic waste and slaughterhouse waste as main substrates. The isolation was initiated by an enrichment using cellulose as a substrate, followed by a procedure described previously (5) using cellulose (5 g/L) as the energy and carbon source. The isolation was performed at the Department of Microbiology, Swedish University of Agricultural Sciences. The genomic DNA was isolated with the DNeasy blood and tissue kit (Qiagen, Inc., Valencia, CA, USA), followed by a cleanup step using the PowerClean Pro DNA cleanup kit (MO BIO Laboratories, Carlsbad, CA, USA). The genomic DNA was then used for constructing a 10-kb SMRTbell library and sequenced using the PacBio RSII single-molecule real-time (SMRT) sequencing platform; the raw reads were assembled with HGAP version 2.0 (Pacific Biosciences, Menlo Park, CA, USA) (6). In total, 137,571 post-filter reads were obtained from two SMRT cells with a mean read length of 11,675 bp. The draft genome of *Clostridium* sp. Bc-iso-3 consists of four contigs with a total length of 4,327,139 bp and an average coverage of 312.97×, among which the max contig length is 4,269,335 bp with a coverage of 315.87×.

The draft genome has a G+C content of 38.2%. In total, four sets of 5S-23S-16S rRNA genes were predicted using RNAmmer (7), and 56 tRNA genes covering all 20 amino acids were predicted using tRNAscan-SE (8). Using PRODIGAL (Prokaryotic Dynamic Programming Gene-finding Algorithm, version 2.6.3), a total of 3,711 coding sequences were predicted (9).

The carbohydrate metabolism enzymes (CAZyme) were annotated using dbCAN (10). In total, 81 CAZymes were found, including one auxiliary activity, 23 carbohydrate-binding modules, nine carbohydrate esterases, 29 glycoside hydrolases, 16 glycosyltransferases, and three polysaccharide lyases.

Accessibility of this genome sequence will further allow a detailed analysis of anaerobic cellulose degradation pathways, as well as a comparative analysis with other bacteria having the ability to degrade cellulose.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number LXSB00000000. The version described in this paper is the first version, LXSB01000000.

REFERENCES

1. **Agbor VB, Cicek N, Sparling R, Berlin A, Levin DB.** 2011. Biomass pretreatment: fundamentals toward application. Biotechnol Adv **29:** 675–685. http://dx.doi.org/10.1016/j.biotechadv.2011.05.005.
2. **Wilson DB.** 2011. Microbial diversity of cellulose hydrolysis. Curr Opin Microbiol **14:**259–263. http://dx.doi.org/10.1016/j.mib.2011.04.004.
3. **Percival Zhang YH, Himmel ME, Mielenz JR.** 2006. Outlook for cellulase improvement: screening and selection strategies. Biotechnol Adv **24:** 452–481. http://dx.doi.org/10.1016/j.biotechadv.2006.03.003.
4. **Azman S, Khadem AF, Van Lier JB, Zeeman G, Plugge CM.** 2015. Presence and role of anaerobic hydrolytic microbes in conversion of ligno-cellulosic biomass for biogas production. Crit Rev Environ Sci Technol **45:**2523–2564. http://dx.doi.org/10.1080/10643389.2015.1053727.
5. **Schnürer A, Schink B, Svensson BH.** 1996. *Clostridium ultunense* sp.

nov., a mesophilic bacterium oxidizing acetate in syntrophic association with a hydrogenotrophic methanogenic bacterium. Int J Syst Bacteriol **46:**1145–1152. http://dx.doi.org/10.1099/00207713-46-4-1145.

6. **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.

7. **Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, 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.

8. **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.

9. **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.

10. **Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y**. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res **40:**W445–W451. http://dx.doi.org/10.1093/nar/gks479.