Complete Genome Sequence of *Bacillus subtilis* Strain WB800N, an Extracellular Protease-Deficient Derivative of Strain 168

Haeyoung Jeong, Da-Eun Jeong, Seung-Hwan Park, Seong Joo Kim, Soo-Keun Choi

ABSTRACT *Bacillus subtilis* WB800N is a genetically engineered variant of *B. subtilis* 168, such that all extracellular proteases are disrupted, which enables WB800N to be widely used for the expression of secretory proteins. Here, we report the 4.2-Mb complete genome sequence of WB800N and present all of the disrupted gene structure.

*Bacillus subtilis* has many advantages as a protein expression host, such as its generally recognized as safe (GRAS) status, easy genetic manipulation, well-known large-scale fermentation process, and protein secretion into media. However, the expression of recombinant secretory proteins in *B. subtilis* has often been unsuccessful due to the degradation of secreted proteins by extracellular proteases (1). *B. subtilis* has eight extracellular proteases, known as NprE, AprE, Epr, Bpr, Mpr, NprB, Vpr, and WprA. To increase the stability of secreted proteins, the eight-extracellular-protease-deficient mutant WB800N was constructed (2) and is commercially available. However, since it was constructed over a long period of time, it is not easy to follow the construction process in order to know the accurate genetic structure of the extracellular protease genes on the genome. Furthermore, the insertion locus of the kanamycin resistance gene is unclear. To clarify this uncertainty, we determined the complete genome sequence of *B. subtilis* WB800N for further strain improvement.

WB800N cells (catalog no. PBS022) were purchased from MoBiTec (Göttingen, Germany) and were grown aerobically in Luria-Bertani (LB) medium at 37°C. Genomic DNA was extracted using a Wizard genomic DNA purification kit from Promega (Wisconsin, USA) according to the manufacturer’s instructions. Library construction and genome sequencing were carried out on a PacBio RS II platform at Chun Lab (Seoul, Republic of Korea) using P6-C4 chemistry. Using the RS_HGAP_Assembly.3 protocol in SMRT Analysis v2.3 (https://www.pacb.com/products-and-services/analytical-software/smrt-analysis/), 73,369 reads totaling 361.44 Mb (81.3× genome coverage and an N50 read length of 7,290 bp) were assembled into two contigs of 4.23 Mb and 1.88 kb. The smaller contig was discarded because it was found to be the PacBio internal control DNA. The chromosomal sequence, corrected through two consecutive rounds of the RS_Resequencing.1 protocol in SMRT Analysis, was circularized using Circlator (3). The final sequence consists of a chromosome of 4,214,174 bp with a G+C content of 43.5%.

Genome annotation was carried out using NCBI’s Prokaryotic Genome Annotation Pipeline (PGAP) v4.6 (4). Compared to the genome sequence of *B. subtilis* strain 168 (GenBank accession no. NC_000964.3) using Cross_match (http://www.phrap.org/), all alignment blocks were nearly identical with each other (0 to 0.01% nucleotide differ-
ence), and they were placed collinearly along the chromosome. Unaligned regions at the boundaries of each block are due to the genetic manipulations in the WB800N strain. Compared to strain 168, the gene structure changes of WB800N are as follows. The nprE, aprE, bpr, vpr, and epr genes were partially deleted without insertion of the antibiotic resistance gene, while wprA and nprB were disrupted by the hygromycin resistance gene and the blasticidin resistance gene, respectively. The entire mpr gene was deleted and replaced by the bleomycin resistance gene. In the middle of the ispA gene, a large DNA fragment containing an rsbRB, a C-terminal fragment of a tetracycline resistance gene, a separated chloramphenicol resistance gene, a kanamycin resistance gene, an ampicillin resistance gene, and an N-terminal fragment of metE were inserted (Fig. 1). The genome information will be useful for further improvement of the strain.

**Data availability.** This genome sequencing project has been deposited in DDBJ/ENA/GenBank under the accession no. CP032310. The version described in this paper is the first version (CP032310.1). Raw sequencing reads are available in NCBI under BioProject accession no. PRJNA490410.

**ACKNOWLEDGMENTS**

This study was supported by the Military Biodefense Laboratory Program (FDC0931823) of the Agency for Defense Development and by the KRRBB Research Initiative Program of the Ministry of Science and ICT, Republic of Korea.
REFERENCES
1. Westers L, Westers H, Quax WJ. 2004. *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. Biochim Biophys Acta 1694:299–310. https://doi.org/10.1016/j.bbamcr.2004.02.011.
2. Nguyen HD, Phan TTP, Schumann W. 2011. Analysis and application of *Bacillus subtilis* sortases to anchor recombinant proteins on the cell wall. AMB Express 1:22. https://doi.org/10.1186/2191-0855-1-22.
3. Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.
4. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44: 6614–6624. https://doi.org/10.1093/nar/gkw569.