Complete Genome Sequence of *Bacillus cereus* NC7401, Which Produces High Levels of the Emetic Toxin Cereulide

Akira Takeno, Akira Okamoto, Keizo Tori, Kenshiro Oshima, Hideki Hirakawa, Hidehiro Toh, Norio Agata, Keiko Yamada, Naotake Ogasawara, Tetsuya Hayashi, Tohru Shimizu, Satoru Kuhara, Masahira Hattori, and Michio Ohta

Department of Bacteriology, Nagoya University Graduate School of Medicine, Showa-ku, Nagoya, Japan; Department of School Nursing and Health Education, Aichi University of Education, Hiroswa, Igaya, Kariya, Aichi, Japan; Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Japan; Department of Plant Genome Research, Kazusa DNA Research Institute, Kisarazu, Chiba, Japan; Medical Institute of Bioregulation, Kyushu University, Higashi-ku, Fukuoka, Japan; Nagoya City Public Health Research Institute, Mizuho-ku, Nagoya, Japan; Graduate School of Biological Science, Nara Institute of Science and Technology, Nara, Japan; Department of Microbiology, School of Medicine, University of Miyazaki, Miyazaki, Japan; Department of Bacteriology, Graduate School of Medical Science, Kanazawa University, Kanazawa, Ishikawa, Japan; Department of Genetic Resources Technology, Faculty of Agriculture, Kyushu University, Higashi-ku, Fukuoka, Japan; and School of Nursing, Sugiyama Jogakuen University, Chikusa-ku, Nagoya, Aichi, Japan

We report the complete and annotated genome sequence of *Bacillus cereus* NC7401, a representative of the strain group that causes emetic-type food poisoning. The emetic toxin, cereulide, is produced by a nonribosomal protein synthesis (NRPS) system that is encoded by a gene cluster on a large resident plasmid, pNCcld.

*Bacillus cereus* is a ubiquitous spore-forming bacterium that has been isolated from people with food-borne illnesses and other infections (4, 9, 11, 19). Food poisoning caused by *B. cereus* is divided into two types according to symptoms: vomit type or diarrhea type. The vomiting type is often life-threatening (6, 13, 17). The toxin cereulide is responsible for the emetic food-borne diseases (1, 2).

*B. cereus*, *Bacillus anthracis*, and *Bacillus thuringiensis* are thought to be descended from a common *Bacillus* ancestor species that adapted to animal hosts (10, 14). The *B. cereus* species includes a large variety of strains that constitute different clusters, one of which is phylogenetically close to the *B. anthracis* group. Our multilocus sequence typing (MLST) study also revealed that most cereulide-producing strains can be allocated to the known sequence type exclusively comprising emetic *B. cereus* strains (18). Here, we report the complete and annotated genome sequence of *B. cereus* strain NC7401 as a representative of this group.

The complete genome sequence of NC7401 was determined by a whole-genome shotgun strategy using the Sanger method. Genomic libraries containing 2-kb and 10-kb inserts were constructed and sequenced. A total of 112,896 sequences were generated from both ends of the genomic clones, giving 10.4-fold coverage from both ends of the genomic clones. Sequence reads were assembled using the Phred-Phrap-Consed program (7, 8). Closing of gaps between contigs was achieved by direct sequencing of the clones. tRNA scans (12) were used to identify tRNA genes. Genome Gambler 1.51 (16), GLIMMER 2.0 (5), and CRITICA (3) were used to identify potential open reading frames (ORFs) larger than 30 amino acids. All predicted proteins were searched against a nonredundant protein database (nr; NCBI) by using BLASTP.

The complete genome of *B. cereus* NC7401 comprises a single circular chromosome of 5,221,581 bp, with a G+C content of 35.6%. The 14 rRNA operons and 104 tRNA genes are mainly located around the putative origin of replication. NC7401 harbors five plasmids (270, 48, 5, 4, and 3 kb in size). The chromosome of NC7401 contains 5,415 protein-coding genes, of which 3,832 are highly conserved among closely related strains in the *Bacillus cereus* sensu stricto group (NC7401, *B. cereus* ATCC 14579, *B. cereus* ATCC 10987, and *B. anthracis* Ames). The average amino acid sequence identities of the 3,832 NC7401 ORFs to the other three strains were 97.0% (ATCC 10987), 96.3% (Ames), and 94.1% (ATCC 14579), indicating that NC7401 is more closely related to *B. anthracis* Ames than *B. cereus* ATCC 14579.

The cereulide biosynthesis gene cluster comprising genes that encode a nonribosomal protein synthesis system was found on the 270-kb plasmid, named pNCcld. The number of assembled sequence reads assembled into pNCcld suggested that one copy of pNCcld is present in a bacterial cell. The genetic structure of pNCcld is almost identical to a 270-kb plasmid, pCER270 (accession number DQ889676), of *B. cereus* strain AH187 (also known as *B. cereus* F4810/72), which was also isolated from a patient with emetic food poisoning (15).

**Nucleotide sequence accession numbers.** The genome sequence has been deposited at DDBJ/EMBL/GenBank under the accession numbers AP007209 (chromosome), AP007210 (pNCcld), AP007211 (pNC1, 48 kb), AP007212 (pNC2, 5 kb), AP007213 (pNC3, 4 kb), and AP007214 (pNC4, 3 kb).

**ACKNOWLEDGMENTS**

We thank K. Furuya, C. Yoshino, H. Inaba, and Y. Hattori (The University of Tokyo) and A. Yamashita, A. Tamura, and N. Itoh (Kitasato University) for their technical assistance. We also thank C. Ito for advice on language and style.

This work was supported by the Research for the Future Program from the Japanese Society for the Promotion of Science (JSPS-RFTF00L01411).

**REFERENCES**