**GENOME ANNOUNCEMENT**

**Genome Sequence of the Thermophile *Bacillus coagulans* Hammer, the Type Strain of the Species**

Fei Su, Fei Tao, Hongzhi Tang, and Ping Xu

State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, People’s Republic of China

Here we announce a 3.0-Mb assembly of the *Bacillus coagulans* Hammer strain, which is the type strain of the species within the genus *Bacillus*. Genomic analyses based on the sequence may provide insights into the phylogeny of the species and help to elucidate characteristics of the poorly studied strains of *Bacillus coagulans*.

*Bacillus coagulans* is known to be a slightly acidophilic, thermostable, and safe species, and it is also an economically important species that is frequently involved in the production of high concentrations of optically pure lactic acid (11, 12), coagulin (7), and other thermostable enzymes. Various morphologies of the cells, spore surfaces, and sporangia have been reported in previous studies; however, these morphologies have complicated the taxonomy of the species (4). *B. coagulans* Hammer (ATCC 7050; DSM1) was first described as the type strain of *B. coagulans* in 1915 by Hammer, who isolated this organism from spoiled canned milk (4). It has been involved in many phylogenetic studies as a reference strain, but little genetic information for this strain is available (3, 9, 14). High-throughput sequence technologies for DNA sequencing have enabled genome-scale phylogenetic studies, which may resolve the incongruences in phylogeny (13). Here we present the whole-genome sequence of this strain; the genomic-level phylogenetic analysis may help us to clarify the diversity of *B. coagulans* strains and elucidate the taxonomic status of the species.

The whole genome of *B. coagulans* Hammer was sequenced using the Illumina HiSeq2000 system; sequencing was performed by the Chinese National Human Genome Center at Shanghai. A total of 10.9 million high-quality-read pairs (2 × 100 bp), which achieved a more-than-200-fold coverage of the genome, were produced for de novo assembly using the VELVET program (version 1.2.03) (16). The protein coding sequences (CDSs) in this genome were predicted using the RAST (rapid annotation using subsystem technology) server (1). Genes encoding tRNAs and rRNAs were identified by using tRNAscan-SE (8) and RNAmmer (6), respectively. In addition, genes were searched against the KEGG (10), UniProt (2), and COG (Clusters of Orthologous Groups) (15) databases to annotate the gene description. To detect the CRISPR clusters, the genome sequence was submitted to the CRISPRFinder program (5).

The draft genome sequence is comprised of 307 assembled contigs with a total length of 3,018,045 bp (N50, 35,029 bp), with a G+C content of 47.2%. About 3,437 CDSs (average length, 750 bp) and 82 tRNA and 10 rRNA genes were predicted. There were 412 subsystems determined by RAST, of which the system of carbohydrate utilization contained 366 proteins. According to the annotation, *B. coagulans* Hammer is predicted to possess complete metabolic pathways, including glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway. However, the absence of the gene encoding xylose isomerase, which initializes the utilization of xylose, prevents *B. coagulans* Hammer from producing lactic acid from low-cost xylose. The genome sequence contains a series of membrane transport systems, including 20 CDSs involved in ATP-binding cassette transporters and 22 CDSs involved in protein translocation, but there is not a complete phosphotransferase system (PTS). The genome also contains about 10 CRISPR-associated proteins, as well as four CRISPR repeats, which may constitute a defense mechanism that allows the host bacterium to survive exposure to foreign genetic elements.

**Nucleotide sequence accession numbers.** The results of this whole-genome shotgun project have been deposited at DDBJ/EMBL/GenBank under the accession number ALAS00000000. The version described in this paper is the first version, ALAS01000000.

**ACKNOWLEDGMENTS**

We acknowledge Huajun Zheng for performance of genome sequencing and analysis at the Chinese National Human Genome Center at Shanghai.

This work was partly supported by the Chinese National Programs for High Technology Research and Development (2011AA02A202 and 2011AA02A207) and the National Natural Science Foundation of China (30821005).

**REFERENCES**


