Running title: Su et al.: Complete genome sequence of Bacillus coagulans

GENOME ANNOUNCEMENTS

Genome Sequence of Thermophilic Strain Bacillus coagulans 2-6, an Efficient Producer of High Optical Purity L-Lactic Acid

Fei Su¹, Bo Yu², Jibin Sun³, Hong-Yu Ou¹, Bo Zhao², Limin Wang², Jiayang Qin¹, Hongzhi Tang¹, Fei Tao¹, Michael Jarek⁵, Maren Scharfe⁵, Cuiqing Ma⁴, Yanhe Ma², and Ping Xu¹*

¹ State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, People’s Republic of China; Institute of Microbiology, Chinese Academy of Sciences, Beijing, People’s Republic of China²; Tianjin Industrial Biotechnology R&D Center, Chinese Academy of Sciences, Tianjin, People’s Republic of China³; State Key Laboratory of Microbial Technology, Shandong University, Jinan, People’s Republic of China⁴, and Helmholtz Center for Infection Research, Inhoffenstrasse 7, Braunschweig, Germany⁵

* Corresponding author. Mailing address for Ping Xu: School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, P. R. China. Phone: 86 21 34206647. Fax: 86 21 34206723. E-mail: pingxu@sjtu.edu.cn.
Abstract

*Bacillus coagulans* 2-6 is an efficient producer of lactic acid. The genome of *B. coagulans* 2-6 owns the smallest genome size among the members of genus *Bacillus* known to date. The frameshift mutation at the start of D-lactate dehydrogenase might be responsible for the production of high optical purity L-lactic acid.

*Bacillus coagulans* was first described in 1915 by Hammer from spoiled canned milk (8). Because of stable high performance in utilization of renewable resource and non-sterilization fermentation under high temperature, the thermophilic *Bacillus coagulans* strains have been suggested to be better producers of lactic acid (11, 13). In addition to the production of lactic acid, *B. coagulans* has also been found to be a source of many other commercially valuable products, such as thermostable enzymes and coagulin, an antimicrobial peptide (6). Compared with other probiotic bacteria such as *Lactobacillus* species, some strains of *B. coagulans* are able to survive in the environment of extremes of heat, acidity of the stomach and bile acids (3). However, little genetic information is known. Here, we represent the genome sequence of a *B. coagulans* strain 2-6, which is an efficient producer of high optical purity L-lactic acid with merits of high carbon-efficiency, less by-product formation, and thermotolerance (13).

The whole genome of *B. coagulans* 2-6 was sequenced using the Illumina GA
system performed by the Helmholtz Center for Infection Research in Germany with a combination of Paired-end library and Mate-pair. And the reads were assembled with the VELVET (14). According to the draft sequence of B. coagulans 36D1 and contigs from different assembly softwares (Edena (5), Euler-SR (2) and SOAPdenovo (7)), the complete genome sequence of 2-6 was completed. Closure of the gaps was finished by Bubble PCR prime walking using the routine Sanger method and edited in Phred/Phrap/Consed (4) package. Finally, Illumina data were used to correct potential base errors and increase consensus quality by mapping the reads to genome. The genome sequence of B. coagulans 2-6 was annotated with NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (12), and functional annotation using Clusters of Orthologous Genes (COG) and KEGG (9).

The genome of B. coagulans 2-6, which is the smallest one in the Bacillus strains by now, is composed of a 3,073,079-bp single circular chromosome with a mean GC content of 47.3% and a 9,910-bp plasmid, whose mean GC content is 38.0%. We identified 2,975 protein-coding sequences (CDS) in the chromosome and 10 CDS in the plasmid. No putative biological functions were predicted in the plasmid. The CDS in the chromosome constitute 79.9% of the genome. The putative biological functions were assigned to 2,332 (78.4%) predicted proteins based on the BLAST (1) results. The frameshift mutation at the start of D-lactate dehydrogenase might be responsible for the production of high optical purity L-lactic acid of strain 2-6 (optical purity > 99%) (13). Only a fragment of pyrophosphokinase in the phosphoketolase pathway was predicted, which suggested the pentose mainly lost in the
transaldolase/transketolase pathway. Compared with the phosphoketolase pathway, the transaldolase/transketolase pathway could produce 1.67 mol of lactic acid per mol pentose, whereas the phosphoketolase pathway only produces 1 mol of lactic acid in addition to 1 mol acetate (10).

Nucleotide sequence accession number. The complete genome sequence of *B. coagulans* 2-6 has been submitted to the GenBank under accession number CP002472.

This work was supported by grants from the National Basic Research Program of China (2007CB707803), Chinese National Programs for High Technology Research and Development (2006AA020102 and 2007AA10Z360), the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-YW-G-005), and the National Natural Science Foundation of China (30900022).

References


