

## Complete Genome Sequence of *Streptococcus thermophilus* Strain ND03<sup>∇</sup>

Zhihong Sun,<sup>2,†</sup> Xia Chen,<sup>2,†</sup> Jicheng Wang,<sup>2</sup> Wenjing Zhao,<sup>2</sup> Yuyu Shao,<sup>2</sup> Lan Wu,<sup>2</sup> Zhemin Zhou,<sup>3</sup>  
Tiansong Sun,<sup>2</sup> Lei Wang,<sup>4</sup> He Meng,<sup>5</sup> Heping Zhang,<sup>2</sup> and Wei Chen<sup>1\*</sup>

State Key Laboratory of Food Science and Technology, School of Food Science and Technology, Jiangnan University, Wuxi 214122, China<sup>1</sup>; Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education, School of Food Science and Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China<sup>2</sup>; The Engineering and Research Center for Microbial Functional Genomics and Detection Technology, Ministry of Education, TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin 300457, China<sup>3</sup>; Tianjin Biochip Corporation, Tianjin 300457, China<sup>4</sup>; and School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China<sup>5</sup>

Received 18 November 2010/Accepted 23 November 2010

***Streptococcus thermophilus* strain ND03 is a Chinese commercial dairy starter used for the manufacture of yogurt. It was isolated from naturally fermented yak milk in Qinghai, China. We present here the complete genome sequence of ND03 and compare it to three other published genomes of *Streptococcus thermophilus* strains.**

*Streptococcus thermophilus* strain ND03 was isolated from naturally fermented yak milk in Qinghai, China (10). It has many excellent processing properties, such as flavor, acidity, viscosity, and water holding. This strain has been implemented in the industrial production of dairy starter cultures by Inner Mongolia Yili Industrial Group Company, Ltd., the largest dairy corporation in China.

Whole-genome sequencing of *S. thermophilus* strain ND03 was performed with a combined strategy of 454 sequencing (9) and Solexa paired-end sequencing technology (1). Genomic libraries containing 3-kb inserts were constructed, and 124,126 paired-end reads and 28,120 single-end reads were generated using the GS FLX system, giving 20.5-fold coverage of the genome. The majority (93.5%) of reads were assembled into seven large scaffolds, including 86 nonredundant contigs, using the 454 Newbler assembler (454 Life Sciences, Branford, CT). A total of 5,647,930 reads (2.5-kb library) were generated to reach a depth of 163-fold coverage with an Illumina Solexa GA IIx (Illumina, San Diego, CA) and mapped to the scaffolds using Burrows-Wheeler alignment (BWA) (7). The gaps between scaffolds were filled by sequencing PCR products using an ABI 3730 capillary sequencer. The genome analysis was performed as described previously (4, 5).

The complete genome sequence of ND03 contains a circular 1,831,957-bp chromosome with a GC content of 39.1%. There are 2,038 genes in total, including 1,919 coding genes, five rRNA operons, and 56 tRNAs in the ND03 genome.

Comparison of the LMG18311 (2), CNRZ1066 (2), LMD-9 (8), and ND03 genomes revealed that they were highly similar, with the exception of 73 encoding genes that are uniquely

present in ND03 but not in the other three strains. Some of the unique genes formed six large insertion islands that were comprised by transposase, glutamate decarboxylase, acetyltransferase, glycosyltransferase, polysaccharide biosynthesis protein, and the exopolysaccharide (EPS) biosynthesis gene cluster.

Similar to other dairy bacteria, *S. thermophilus* is able to synthesize EPSs that lead to an improvement in the viscosity and texture of yogurt (3). The ND03 genome carries a unique 23.4-kb EPS gene cluster (STND\_1010 to STND\_1035), which contains 10 EPS-related genes and six intact or truncated insertions (IS). Four of the EPS-related genes in the cluster, *epsA*, *epsB*, *epsC*, and *epsD*, were conserved between all four genomes in comparisons. These genes are involved in the regulation, polymerization, and chain length determination and export of the EPS. The remaining six genes (*epsE*, *epsF*, *epsG*, *epsI*, *epsJ*, and *epsP*) in the EPS gene cluster were uniquely present in ND03 and regarded as the key enzymes to determine the formation of a special EPS (6).

Interestingly, six copies of IS that belong to the IS3, IS6, and ISL3 families were found in the EPS gene cluster. This increases the possibility that the unique EPS genes were imported by the transposition of these IS. Similar situations were discovered in many other polysaccharide gene clusters (11). Among these IS, two copies of ISL3 and one copy of IS3 were truncated by frameshift. That indicates that these three IS have lost their transposition capability and may have been in this gene cluster for a long time (11).

**Nucleotide sequence accession number.** The sequence and annotation of the *Streptococcus thermophilus* ND03 genome is available from GenBank under accession number CP002340.

\* Corresponding authors. Mailing address: State Key Laboratory of Food Science and Technology, School of Food Science and Technology, Jiangnan University, 800 Lihu Avenue, Wuxi 214122, China. Phone and fax: 86-510-85912155. E-mail: weichen@jiangnan.edu.cn.

† Both authors contributed equally to this work.

∇ Published ahead of print on 3 December 2010.

This research was supported by National Natural Science Foundation of China (grant no. 30760156, 30800861, 30860219, and 31025019), Hi-Tech Research and Development Program of China (863 Program) (2010AA10Z302), the Earmarked Fund for Modern Agro-Industry Technology Research System, the Prophase Research Program of the 973 Project of China (2010CB134502), National Key

Technology R&D Program (2009BAD1B01), and the Innovation Team Development of the Ministry of Education of China (IRT0967).

## REFERENCES

1. Bentley, D. R., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**:53–59.
2. Bolotin, A., et al. 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat. Biotechnol.* **22**:1554–1558.
3. Duboc, P., and B. Mollet. 2001. Applications of exopolysaccharides in the dairy industry. *Int. Dairy J.* **11**:759–768.
4. Feng, L., et al. 2008. A recalibrated molecular clock and independent origins for the cholera pandemic clones. *PLoS One* **3**:e4053.
5. Ferenci, T., et al. 2009. Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12. *J. Bacteriol.* **191**:4025–4029.
6. Jolly, L., and F. Stingle. 2001. Molecular organization and functionality of exopolysaccharide gene clusters in lactic acid bacteria. *Int. Dairy J.* **11**:733–745.
7. Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.
8. Makarova, K., et al. 2006. Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **103**:15611–15616.
9. Margulies, M., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
10. Sun, Z., et al. 2010. Identification and characterization of the dominant lactic acid bacteria from kurut: the naturally fermented yak milk in Qinghai, China. *J. Gen. Appl. Microbiol.* **56**:1–10.
11. Wang, L., and P. R. Reeves. 1998. Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect. Immun.* **66**:3545–3551.