J. Dairy Sci. 101:1–7
https://doi.org/10.3168/jds.2017-13534

Short communication: Genomic and phenotypic analyses of exopolysaccharides produced by *Streptococcus thermophilus* KLDS SM

Bailiang Li,*[1] Xiuyun Ding,*[1] Smith Etareri Evivie,*† Da Jin,* Yueyue Meng,* Guicheng Huo,*[2] and Fei Liu*
*Key Laboratory of Dairy Science, Ministry of Education, College of Food Science, Northeast Agricultural University, Harbin, 150030, China
†Food Science and Nutrition Unit, Department of Animal Science, Faculty of Agriculture, University of Benin, Benin City, PMB 1154, Nigeria

ABSTRACT

Streptococcus thermophilus plays important roles in the dairy industry. *Streptococcus thermophilus* KLDS SM could produce a high amount of exopolysaccharides (EPS). To understand the possible link between the genotype and the phenotype regarding EPS, the complete genome of *S. thermophilus* KLDS SM was sequenced and investigated in silico for genes related to carbohydrate fermentation, nucleotide sugars synthesis, and EPS gene cluster. We found that *S. thermophilus* KLDS SM is able to ferment sucrose, mannose, glucose, galactose, and lactose from the genomic research, which was confirmed by API 50 CH (bioMérieux, Marcy l'Etoile, France). The genetic analysis of nucleotide sugars and EPS cluster revealed that the EPS produced by this strain are composed of galactose and glucose, in accordance with the biochemical result. Furthermore, differences in the molecular mass of EPS from *S. thermophilus* KLDS SM cultivated under different carbon sources were correlated with the transcription levels of the genes encoding chain length determination protein and glycosyltransferase. Our findings provide a better understanding of the link between the genetic elements and the chemical conformation of EPS and a theoretical basis for producing tailor-made EPS through genetic and metabolic engineering approaches.
Key words: *Streptococcus thermophilus*, genome sequence, monosaccharide composition, molecular mass

Short Communication

Streptococcus thermophilus is a well-known, nonpathogenic lactic acid bacterium (**LAB**). Combined with *Lactobacillus delbrueckii* ssp. *bulgaricus*, it is traditionally used as a starter for fermented dairy food products, such as yogurt, cheese, and ice cream (Liu et al., 2009; Settachaimongkon et al., 2014). Some previous studies suggested that *S. thermophilus* plays a vital role in treating lactose intolerance, antioxidant production, stimulation of intestinal immune responses, and relieving some cancers (Iyer et al., 2010).

Several *S. thermophilus* strains can produce exopolysaccharides (**EPS**), which can improve rheological and sensory attributes of fermented food products (Caggianiello et al., 2016). Moreover, some EPS from LAB have been shown to possess probiotic properties, such as antioxidant and antitumoral activities, reducing cholesterol, regulating the immune system, and providing substrates for the intestinal flora, which are related to the chemical conformation of EPS (Caggianiello et al., 2016; Cui et al., 2017). However, few reports offer scientific confirmation of the link between the genetic elements and the chemical conformation of EPS.

Streptococcus thermophilus KLDS SM was isolated from naturally fermented yogurt in Inner Mongolia, and it was shown to produce high levels of EPS (B. Li, F. Liu, and G. Huo, unpublished data). Based on this characteristic, *S. thermophilus* KLDS SM was selected to further clarify the possible link between the genotype and the phenotype regarding EPS. We analyzed the genetic elements involved in EPS biosynthesis to determine the structure of EPS and the transcriptional level of EPS-related genes under different carbon sources.

The genomic DNA of *S. thermophilus* KLDS SM was extracted by the DNeasy Tissue kit (Qiagen, Hilden, Germany). The whole genome sequencing was performed using a combined strategy of Illumina Hiseq 2500 sequencing (insert size of 500 bp; Illumina, San Diego, CA) and Pacific Biosciences (**PacBio**) RSII sequencing (20,000 bp template library; PacBio, Menlo Park, CA) technologies. A total of 63,855 subreads and a total of 401 Mb of clean pair-end reads were obtained, respectively. Subsequently, PacBio RSII reads were de novo assembled into a circular contig with an average genomic coverage of 240 folds using the RS hierarchical

Received July 20, 2017.
Accepted August 26, 2017.
[1]These authors contributed equally to this work.
[2]Corresponding author: guichenghuo@126.com

genome assembly progress protocol (Chin et al., 2013). Pair-end reads were then used to correct the single base errors of PacBio RSII reads by SOAPsnp v1.05 (Li et al., 2009). Genome annotation was performed using NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/books/NBK174280). The circular genomic map was constructed using CGView Server (Grant and Stothard, 2008). Functional annotation of protein-coding genes (**CDS**) was performed with the Kyoto Encyclopedia of Genes and Genomes (**KEGG**) database (http://www.genome.jp/kegg/). Family distribution, conserved domain, and model structure of glycosyltransferase (**GT**) were analyzed using the carbohydrate-active enzymes database (http://www.cazy.org/), Conserved Domain Database (https://www.ncbi.nlm.nih.gov/cdd/?term=), and SWISS-MODEL server (https://swissmodel.expasy.org/), respectively.

The profiles of carbohydrate fermentation were analyzed by API 50 CH test strip and API CHL medium (bioMérieux, Marcy l'Etoile, France) according to the manufacturer's instructions. Glucose, lactose, or sucrose was used in the chemically defined medium (Letort and Juillard, 2001) as the only carbon source at the final concentration of 1% (wt/vol). The EPS samples from chemically defined medium were extracted and then purified by DEAE-Sepharose Fast Flow (GE Healthcare, Pittsburgh, PA) and Sepharose CL-6B (GE Healthcare) following the method described by Ren et al., (2016). The monosaccharide composition and the molecular mass distribution of EPS were performed using high-performance anion exchange chromatography (ICS-3000, Dionex, Sunnyvale, CA) and high-performance size-exclusion chromatography (1260, Agilent, Santa Clara, CA) as previously described (Shao et al., 2014). The RNA isolation, reverse transcription, and real-time quantitative PCR (**RT-qPCR**) were implemented according to a previous protocol (Li et al., 2016). The specific primers used for the RT-qPCR are listed in Supplemental Table S1 (https://doi.org/10.3168/jds.2017-13534).

The genome of *S. thermophilus* KLDS SM consists of a circular chromosome (1,856,787 bp) with G + C content of 39.08% (Figure 1A). A total of 1,950 genes were predicted, including 1,732 CDS, 129 pseudogenes, 18 rRNA genes, 67 transfer RNA genes, and 4 noncoding RNA genes (Supplemental Table S2; https://doi.org/10.3168/jds.2017-13534). As shown in Figure 1B, 1,046 CDS were annotated into KEGG database; moreover, the highest number of 150 genes in the functional group associated with carbohydrate metabolism was found in the genome, suggesting that *S. thermophilus* KLDS SM may have advantage in carbohydrate utilization and EPS biosynthesis. Additionally, the genome neighbor report (Supplemental Table S3; https://doi.org/10.3168/jds.2017-13534) revealed that *S. thermophilus* KLDS SM had 99.5673% symmetric identity with *S. thermophilus* ASCC 1275, which was reported to give a high EPS yield (Zisu and Shah, 2003).

Exopolysaccharide biosynthetic pathway involves the sugar uptake system, nucleotide sugar synthesis, polysaccharide synthesis, and export of the EPS (Laws et al., 2001; Cui et al., 2017). Genetic elements with EPS biosynthetic pathway were mined in the genome of *S. thermophilus* KLDS SM from these aspects. The most efficient sugar transport is the phosphoenolpyruvate-phosphotransferase system (**PTS**), which is composed of histidine-containing phosphoprotein, phosphoenolpyruvate-dependent phosphotransferase and sugar-specific permease enzyme (Laws et al., 2001). The genome of *S. thermophilus* KLDS SM harbors the histidine-containing phosphoprotein (A9497_02385) and phosphoenolpyruvate-dependent phosphotransferase (A9497_02380); in addition, genes responsible for sucrose and mannose PTS transporter sugar-specific permease enzyme were found in the genome (Supplemental Table S4; https://doi.org/10.3168/jds.2017-13534). Thus, we suggested that sucrose and mannose are the only 2 sugars that may be transported by PTS. Genes encoding lactose/galactose permease (A9497_02960) and glucose permease (A9497_00655) were identified in the genome. The ATP-binding cassette (**ABC**) transporter is not available to transport sugars into the cytoplasm of *S. thermophilus* KLDS SM, because it only equips 2 genes encoding ABC transporter (Supplemental Table S4). We found that *S. thermophilus* KLDS SM is able to use sucrose, mannose, glucose, galactose, and lactose from the genomic analysis. Carbohydrate fermentation results from the API 50 CH galleries also showed that *S. thermophilus* KLDS SM could metabolize these 5 sugars.

Once sugars are transported into the cytoplasm, they will be metabolized to the nucleotide sugars by various pathways, which act as active precursors to contribute to the EPS structure. As shown in Figure 2A, *S. thermophilus* SM could hydrolyze lactose into glucose and galactose by β-galactosidase (A9497_02955), then galactose is converted to the glucose-1-phosphate and the uridine diphosphate (**UDP**)-galactose the by the Leloir pathway (A9497_02965, A9497_02980, and A9497_02975). Glucose is phosphorylated by glucokinase (A9497_00055) to glucose-6-phosphate, which is further mutated to the glucose-1-phosphate by phosphoglucomutase (A9497_00345). Glucose-1-phosphate, as the vital intermediate product, could be transformed to the UDP-glucose by UDP-glucose pyrophosphorylase (A9497_05025) or to the deoxythymidine diphosphate (**dTDP**)-rhamnose using a series of enzymes encoded by A9497_02280, A9497_02270,