Genomic and phenotypic analyses of exopolysaccharide biosynthesis in Streptococcus thermophilus S-3

Zhi-Qiang Xiong, Ling-Hui Kong, Phoency F.-H. Lai, Yong-Jun Xia, Ji-Chao Liu, Quan-Yang Li, and Lian-Zhong Ai

1Shanghai Engineering Research Center of Food Microbiology, School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China
2Beijing SANYUAN Foods Co. Ltd., Beijing 100076, China
3College of Light Industry and Food Engineering, Guangxi University, Nanning 530004, China

ABSTRACT

Streptococcus thermophilus, one of the most important industrial lactic acid bacteria, is widely used as a starter culture in the dairy industry. Streptococcus thermophilus S-3 isolated from Chinese traditional dairy products has shown great potential for the production of larger amounts of exopolysaccharides (EPS), which significantly affect the organoleptic properties of fermented milk products. To understand the relationship between the genotype and phenotype of S. thermophilus S-3 in terms of EPS biosynthesis, its genome of strain S-3 was sequenced and the genes related to carbohydrate utilization, nucleotide sugars synthesis, and EPS biosynthesis were investigated. The genomic analysis revealed that S. thermophilus S-3 can use sucrose, mannose, glucose, galactose, and lactose. Phenotypic analysis showed that S-3 prefers fermenting lactose to fermenting glucose or galactose. The genetic analysis of nucleotide sugars and EPS biosynthesis revealed that S-3 can synthesize uridine diphosphate (UDP)-glucose, deoxythymidine diphosphate-glucose, deoxythymidine diphosphate-rhamnose, UDP-galactose, UDP-N-acetylgalactosamine, and UDP-N-acetylglucosamine. A high yield of EPS from S-3 cultivated with lactose rather than glucose as the carbon source was correlated with high transcriptional levels of the genes associated with metabolism of these nucleotide sugars and EPS biosynthesis. Our results provide a better understanding of EPS biosynthesis in S. thermophilus and can facilitate enhanced EPS production by lactic acid bacteria fermentation via genetic and metabolic engineering approaches.

Key words: exopolysaccharide biosynthesis, genome analysis, nucleotide sugar, Streptococcus thermophilus

INTRODUCTION

Among the industrial lactic acid bacteria (LAB), Streptococcus thermophilus is considered a generally regarded as safe (GRAS) microorganism with non-pathogenic, fermentative, and facultative anaerobic traits (Vaningelgem et al., 2004; Hols et al., 2005; De Vuyst et al., 2011; Wu et al., 2014). As one of the most important homo-fermentative thermophilic bacteria, S. thermophilus is widely used in the production of fermented dairy products such as yogurt and cheese (Liu et al., 2009; Iyer et al., 2010). The characteristics of S. thermophilus relevant to dairy products include rapid acidification, bacteriocin synthesis, lactose utilization, and exopolysaccharide (EPS) synthesis (Rasmussen et al., 2008; Wu et al., 2014).

Exopolysaccharide can be classified into homopolysaccharides composed of a single type of sugar and heteropolysaccharides with repeating units of various sugars (Cui et al., 2016). Although the physiological roles of EPS in LAB are yet to be completely elucidated, the recognized functions of EPS include (1) protecting LAB from extreme environments such as bile and osmotic stresses, (2) participating in LAB adhesion/recognition and biofilm formation, and (3) conferring health benefits such as immune-stimulatory effects (Caggianiello et al., 2016). Exopolysaccharides are widely used for food thickening, stability, emulsification, gelation, and water holding (Freitas et al., 2011; Marette and Picard-Deland, 2014). In the dairy industry, EPS have been shown to act as natural thickeners that prevent syneresis and improve the viscosity, texture, and mouthfeel of dairy products (Cui et al., 2017). Accordingly, EPS biosynthesis is recognized as one of the most important properties of S. thermophilus.

In general, EPS biosynthesis is associated with cell growth in S. thermophilus (Cui et al., 2017). Exopolysaccharides are generally produced by S. thermophilus strains at relatively low levels (Svensson et al., 2005; Li et al., 2016). A large number of studies have inves-
tigated the improvement of EPS yields by optimizing fermentation conditions (Zisu and Shah, 2003; Cui et al., 2017; Li et al., 2018). Cultivation conditions (e.g., carbon and nitrogen sources) can significantly affect EPS production (Li et al., 2016). For example, Li et al. (2016) found a large difference in EPS yield with different carbon sources including sucrose, glucose, fructose, lactose, and galactose, and the amount of EPS reached up to 120 mg/L using 8% sucrose in S. thermophilus 05–34. Most EPS of S. thermophilus are heteropolysaccharides that predominantly consist of galactose and glucose in various ratios (Li et al., 2018; Xu et al., 2018; Zhang et al., 2018), accompanied with low levels of acetylgalactosamine, rhamnose, fucose, and acetylated galactose moieties. Brief descriptions of EPS biosynthesis in S. thermophilus have been reported (Stingele et al., 1996; Cui et al., 2017). At the genetic level, the synthesis of EPS by the eps gene cluster is a complex process in which nucleotide sugars are used as the building blocks. The eps cluster in S. thermophilus is around 15 to 36 kb and contains the genes responsible for the regulation, determination of chain length, synthesis, polymerization, and exportation of the repeating sugar units concerned (Cui et al., 2017).

Wild-type S. thermophilus S-3 was isolated from Chinese traditional dairy products in our laboratory. Because of the higher yield of EPS, yogurt fermented by S-3 shows a promisingly high viscosity and thick mouth feel (Xu et al., 2018). A structural analysis of the EPS produced by S-3 revealed that it was newly found (Xu et al., 2018). Few studies have been done about the effect and mechanism of sugars on EPS biosynthesis in S. thermophilus. Accordingly, in this study, the genome of S-3 was sequenced for elucidation of EPS biosynthesis. The effects of lactose and glucose on EPS biosynthesis in S-3 were analyzed based on transcriptional changes of the genes involved in nucleotide sugar synthesis and the eps gene cluster.

**MATERIALS AND METHODS**

**Bacterial Strain and Culture Conditions**

A typical individual colony of S. thermophilus S-3 was anaerobically incubated in M17 agar medium with 2% lactose at 37°C for 24 h. To determine the ability of S-3 to use various sugars, overnight cultures grown in M17 medium with 2% lactose were inoculated by 1:100 dilution into M17 medium supplemented with the appropriate sugar as the sole carbon and energy source, and cell growth was determined by optical density at 600 nm (van den Bogaard et al., 2004). The Baranyi model (Baranyi and Roberts, 1994) was used to estimate the growth parameters maximum growth rate ($\mu_{max}$) and lag phase duration ($\lambda$) in S-3. The culture was inoculated into M17 medium containing 2% lactose, glucose, or galactose at 37°C for 24 h under anaerobic conditions for EPS production.

**Analysis of S. thermophilus S-3 Based on Genome Data**

The genomic DNA of S-3 was extracted using a DNeasy Tissue Kit (Qiagen, Hilden, Germany) and the library was prepared using a TruSeq DNA Sample Prep Kit (Illumina, San Diego, CA) according to the manufacturer’s instructions. The quantity and quality of DNA was measured using a Nanodrop spectrophotometer (ThermoFisher Scientific, Wilmington, DE) and agarose gel electrophoresis, respectively. Genome data were generated by an Illumina Miseq platform and assembled using SOAPdenovo software (Xie et al., 2014). The circular genomic map was constructed by the CGView Server (Grant and Stothard, 2008). Putative protein-coding sequences (CDS) were identified using the Glimmer system (Delcher et al., 2007). Functional annotation of CDS was performed by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg) and the Clusters of Orthologous Groups of proteins (COG) database (https://www.ncbi.nlm.nih.gov/COG/). The eps gene cluster was identified based on BLASTN (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch) results with the genome data of S. thermophilus LMD-9 (NC_008532). The respective sequences were extracted from the genome data and compared using Vector NTI Advance software (v.11.5, Invitrogen, Thermo Fisher Scientific, Waltham, MA).

**RNA Preparation and Real-Time Quantitative PCR Analysis**

The S-3 cells from the M17 medium with 2% lactose or glucose were collected at 24 h, and total RNA was isolated using an RNA isolation kit (Axygen, Corning Inc., Jiangsu, China). Genome DNA was removed by digestion with DNase I (Takara Biomedical Technology Co. Ltd., Beijing, China) and verified by PCR analysis with the RNA as the template. The quantity and integrity of RNA was determined by both Nanodrop spectrophotometer and agarose gel electrophoresis. The cDNA prepared with a PrimeScript RT Reagent Kit (Takara) was used as a template for real-time quantitative PCR with a Lightcycle 96 RT-PCR system (Roche Diagnostics, Basel, Switzerland) using an SYBR Premix Ex TaqII kit (Takara). The real-time quantitative PCR primers are listed in Table 1. Each 20-µL PCR reaction contained 10 µL of real-time PCR mix, 0.5 µL
of template cDNA, and 0.5 μL of 20 μM primers. The PCR reaction programs were as follows: pre-heating at 95°C for 5 min, 40 cycles of 95°C for 30 s, 95°C for 5 s, 60°C for 30 s, and a final extension at 72°C for 5 min. The data for each gene were analyzed and normalized using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) with recA encoding recombinase A as the internal control (Song et al., 2018). All data were verified in 3 independent experiments.

### Extraction and Determination of EPS

Crude EPS was extracted according to the modified method described by Ren et al. (2016). The fermentation broth was successively heated in boiling water for 10 min to inactivate enzymes, cooled to room temperature, centrifuged (10,000 × g, 20 min, and 4°C) for removal of cells and coagulated proteins, and the supernatant was collected. Trichloroacetic acid was then added to the supernatant at a final concentration of 4% (wt/vol) and held at 4°C for 8 h. The protein precipitate was removed by centrifugation, and the supernatant was dialyzed in a dialysis bag (molecular weight cut off of 1 kDa) for 3 d. The insoluble precipitate was removed by centrifugation. The EPS in the final supernatant was collected and its total carbohydrate content was measured using a phenol-sulfuric acid assay (Dubois et al., 1956).

### RESULTS AND DISCUSSION

#### Genome Sequencing of S. thermophilus S-3

The genomes of an increasing number of S. thermophilus strains have been sequenced to study various genetic traits such as virulence potential and the mechanisms for obtaining a high EPS yield (Bolotin et al., 2004; Wu et al., 2014; Li et al., 2018). In this study, to provide insight into the biosynthesis of newly described EPS in S-3, Illumina HiSeq sequencing was used to sequence the genome of S-3. The draft genome of S-3 was distributed into 60 scaffolds totaling 1,743,641 bp with an average GC content of 38.92%, which consisted of one linear chromosome with 3 rRNA operons, 36 tRNA genes, and 1,931 CDS (Figure 1A). For the CDS, 1,155 and 1,483 proteins could be assigned to the KEGG and COG databases, respectively. Our result is similar to those for previously sequenced strains of S. thermophilus such as ASCC 1275 (Wu et al., 2014), KLD SM (Li et al., 2018), and CNRZ1066 and LMG18311 (de Vin et al., 2005) isolated from various regions for which the genome sizes are ~1.8 M with GC content of ~39%. According to COG annotation, the highest number of genes in the genome of S-3 (180) was in the functional group associated with AA transport and metabolism, suggesting that these genes might be closely related to the adaptation of S. thermophilus to a milky environment and nutrients such as milk proteins (Wu et al., 2014). In addition, 91 genes were discovered in the functional groups associated with carbohydrate transport and metabolism, implying that strain S-3 might have an advantage in carbohydrate utilization and EPS biosynthesis.

### Table 1. Real-time quantitative PCR primers used in this study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer (5’ → 3’)</th>
<th>Reverse primer (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>epsA</td>
<td>AGGAACCAGTGCTCATATTC</td>
<td>CTCGGTATTTTCCTGTGTTGG</td>
</tr>
<tr>
<td>epsC</td>
<td>TCACGTACCAACCAACACC</td>
<td>AGATGTCACAACACCCTGGAAG</td>
</tr>
<tr>
<td>epsE</td>
<td>ACTTCTCTCTGGAAACTGTC</td>
<td>TCCAGTGCCACTACCGATAG</td>
</tr>
<tr>
<td>epsI</td>
<td>CCAATAGAGCAGAAGGCAAAG</td>
<td>5GGTTATCTTGTGTTATCTGAC</td>
</tr>
<tr>
<td>epsII</td>
<td>AAACATCAACCCAGGACCC</td>
<td>ATAAAAAGTCAACATTTTCAAACAGG</td>
</tr>
<tr>
<td>lacZ</td>
<td>CCACCTCATCCTCAAGAAAATGCTCAATATTCACCAACCCA</td>
<td>CCTCGAGGACCACTATTGTGTCG</td>
</tr>
<tr>
<td>lact</td>
<td>CCATGCTGGTCTCTATCGATGT</td>
<td>5AGATGTCACAACACCCTGGAAG</td>
</tr>
<tr>
<td>galM</td>
<td>GAGGAAGTCTGTTACATGCG</td>
<td>TCCGAGGACCACTATTGTGTCG</td>
</tr>
<tr>
<td>galE</td>
<td>ACCACAAAATCTATCAACACCA</td>
<td>TCAACACTGACACCATCACGG</td>
</tr>
<tr>
<td>galT</td>
<td>TTGGGAGAAGGTTGACTGATG</td>
<td>AGATCCATGATGTCAGCACC</td>
</tr>
<tr>
<td>galK</td>
<td>AGAGCTGGGCTCATACTGATG</td>
<td>CAGGGCTTCAATTTAATGTC</td>
</tr>
<tr>
<td>galU</td>
<td>TCAGGGTAAAGGGCCAAATGG</td>
<td>CAGAAAATCTAGGTTAAAGAGG</td>
</tr>
<tr>
<td>gpr</td>
<td>CTGCTGGAATGTGTTGTCGG</td>
<td>TCCTGGTTTTTTGCGGGTCCAG</td>
</tr>
<tr>
<td>glnl</td>
<td>AGGATTTGAGTGTACGGCG</td>
<td>5CAAATCACCATTGTCTAGCAG</td>
</tr>
<tr>
<td>recA</td>
<td>TGATTGACTCTGGTGCACTG</td>
<td>AGAAGTTTACCGCATGCTTG</td>
</tr>
</tbody>
</table>

Because it is very important to guard against the failure of milk fermentation in both dairy and starter culture industries, this may provide an advantage when using S-3 for the industrial manufacture of dairy products.
Figure 1. Circular genome map of *Streptococcus thermophilus* S-3 (A) and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation of phosphotransferase system (B). From outside to inside, circles (1, 2) show gene locations including protein-coding genes (CDS), rRNA, tRNA, and other genes on positive and negative chains. Circles (3, 4) show GC content and GC skew, respectively. The red box in the KEGG (http://www.genome.jp/kegg) map indicates the gene annotated in the genome of strain S-3. ORF = the open reading frame in the DNA sequence. GC = guanine and cytosine.
Sugar Uptake in EPS Biosynthesis Based on Genome Analysis of S. thermophilus S-3

Biosynthesis of EPS involves sugar uptake into the cytoplasm, nucleotide sugar synthesis, polysaccharide synthesis, and export (Cui et al., 2017). The phosphoenolpyruvate phosphotransferase system (PEP-PTS) is the most common sugar transport system in bacteria. Because the EPS in *S. thermophilus* are mainly composed of galactose and glucose (Li et al., 2018; Xu et al., 2018; Zhang et al., 2018), genes associated with galactose- and glucose-related transporter and metabolism were mined from the genome of S-3 in this study. It was interesting to find genes responsible for glucose, β-glucosides, trehalose, sucrose, and fructose-specific PTS systems (Figure 1B). This finding suggests that, in addition to sucrose and fructose, glucose, β-glucosides, and trehalose may also be transported by the PEP-PTS in S-3. This result contrasts with the finding of van den Bogaard et al. (2004) that only sucrose and fructose are transported through the PEP-PTS in *S. thermophilus*.

 Sugars without a specialized PEP-PTS are transported into the cytoplasm through non-PEP-PTS systems such as primary and secondary transport systems (Cui et al., 2016). In *S. thermophilus*, sugars such as lactose are mainly transported by secondary transport system. LacS, the gene that encodes lactose permease (a secondary transport system responsible for the import and export of lactose and galactose) was found in the S-3 genome, indicating that S-3 can transport lactose and galactose from the environment. The metabolism of lactose and galactose in *S. thermophilus* involves the Leloir pathway encoded by *gal-lac* operons, which are induced by either lactose or galactose (Anbukkarasi et al., 2014). The *lac* operon includes *lacS* and *lacZ* encoding β-galactosidase for hydrolyzing lactose into glucose and galactose. The *gal* operon encoding galactokinase (*GalK*), galactose-1-phosphate uridylyltransferase (*GalT*), UDP-glucose-4-epimerase (*GalE*), and galactosemutarotase (*GalM*) is responsible for galactose metabolism in *S. thermophilus* (van den Bogaard et al., 2004). The operon-specific regulator *galR* can activate the transcription of *gal-lac* operons and repress their expression (Vaughan et al., 2001). All genes of the *gal-lac* operons were found in the genome of S-3, which suggests that S-3 may use lactose and galactose as carbon sources.

 The ability of S-3 to use lactose, galactose, and glucose was tested in this study. The growth curves of S-3 grown for 24 h on M17 medium supplemented with different carbohydrates as the sole carbon and energy source are shown in Figure 2. The results suggest that S-3 can metabolize lactose, galactose, and glucose as the sole carbon and energy source. They also suggest that S-3 has a galactose-positive (Gal+) phenotype. The main growth parameters λ and μₘₐₓ were estimated by applying the Baranyi growth model. The λ was significantly higher on galactose and glucose than on lactose, and μₘₐₓ was significantly lower on galactose and glucose than on lactose (Table 2), suggesting that cell growth of S-3 is more efficient on lactose than on galactose or glucose. A higher concentration (4%) of galactose also reduced cell growth. The results are similar to those for *S. thermophilus* strains isolated from dairy products in Italy by Tarrah et al. (2018). Most, *S. thermophilus* strains prefer lactose to galactose or glucose as a major carbon and energy source (Anbukkarasi et al., 2014) because they have a galactose-negative (Gal−) phenotype that cannot metabolize galactose. Accordingly, they only use the glucose moiety of lactose and release the residual galactose moiety into the extracellular environment (Vaughan et al., 2001; de Vin et al., 2005).

 It has been reported that the difference between Gal+ and Gal− strains of *S. thermophilus* is mutations of *gal* operon (e.g., the *gal* promoter; Anbukkarasi et al., 2014). The *gal* operon in S-3 was compared with that of a Gal− strain, *S. thermophilus* AR333 (Ren et al., 2016). Various genes of the *gal* operon were identical and showed no substitutions at all in these 2 strains. Compared with AR333, only a single nucleotide substitution was present in the *galK* promoter, a T to C substitution at the −9 box in S-3 (data not shown). This suggests that this single point mutation of the *galK* promoter may largely be responsible for the Gal+ phenotype of S-3. Giaretta et al. (2018) also compared the *gal* promoters of Gal+ and Gal− strains and found mutations at various positions in the *gal* operon intergenic regions. These mutations result in insufficient expression of the *gal* operon genes in the Gal− strains.

Nucleotide Sugar Synthesis in EPS Biosynthesis Based on Genome Analysis of S. thermophilus S-3

 Sugars are metabolized to nucleotide sugars, which are precursors of EPS, by various pathways in the cytoplasm. All genes involved in nucleotide sugar synthesis were identified in the genome of S-3 (Figure 3). As shown in Figure 3A, lactose can be hydrolyzed into glucose and galactose by β-galactosidase encoded by *lacZ*. Galactose is then transformed to glucose-1-phosphate and uridine diphosphate (UDP)-galactose by a series of enzymes including *GalK*, *GalT*, and *GalM*, whereas *GalE* is responsible for the interconversion of UDP-glucose and UDP-galactose. Glucose is phosphorylated to glucose-6-phosphate under the catalysis of glucokinase and further mutated by phosphoglucomutase to glucose-1-phosphate, a key intermediate for nucleotide sugar synthesis. Glucose-1-phosphate can be converted...
to UDP-glucose by UDP-glucose pyrophosphorylase or to deoxythymidine diphosphate (dTDP)-rhamnose by sequential catalysis involving dTDP-glucose pyrophosphorylase, dTDP-glucose 4,6-dehydratase, dTDP-4-dehydrorhamnose 3,5-epimerase, and dTDP-4-dehydrorhamnose reductase. Three copies of galE were found in the genome of S-3, allowing flexibility for the conversion between UDP-glucose and UDP-galactose as reported for other *S. thermophilus* strains (Li et al., 2018). Fructose is transported into the cytoplasm by PEP-PTS and phosphorylated to fructose-1-phosphate by fructokinase, which is then converted to fructose-6-phosphate by phosphomutase. Fructose-6-phosphate is then converted to UDP-*N*-acetylgalactosamine under the catalysis of glucosamine-6-phosphate deaminase, phosphoglucosamine mutase, glucosamine-1-phosphate N-acetyltransferase, N-acetylgalcosamine-1-phosphate uridytransferase, and GalE. Therefore, at least 6 nucleotide sugars (UDP-glucose, dTDP-glucose, dTDP-rhamnose, UDP-galactose, UDP-*N*-acetylglactosamine, and UDP-*N*-acyetylglucosamine) can be formed in S-3 via in silico KEGG analysis. These nucleotide sugars may be assembled into the repeating units of EPS.

**EPS Synthesis and Export Based on Genome Analysis of S. thermophilus S-3**

Biosynthesis of EPS is regulated and determined by the *eps* biosynthetic gene cluster located on the plasmid or chromosome in LAB (Cui et al., 2016). In this study, all 13 *eps* genes were identified in the genome of S-3 (*espA* to *espM*; Figure 3B). This *eps* gene cluster is involved in the regulation of EPS biosynthesis (*epsA* and *epsB*), chain length determination of EPS (*epsC* and *epsD*), biosynthesis of the repeating sugar units (*epsE*, *epsF*, *epsG*, *epsH*, and *epsI*), and EPS polymerization and export (*epsJ*, *epsK*, *epsL*, and *epsM*), which was almost identical to the *eps* gene cluster in *S. thermophilus* Sfi6 (Stingele et al., 1996). The *epsE* gene encodes galactosyltransferase, the priming glycosyltransferase (GT) for transferring sugar to the undecaprenylphosphate lipid carrier as the first step in the synthesis of the repeating unit (Stingele et al., 1996). Minic et al. (2007) reported that EpsE (phosphoglycosyltransferase), which initiates exopolysaccharide synthesis, is controlled by EpsD (tyrosine kinase) in *S. thermophilus*.

The *epsF*, *epsG*, and *epsI* genes are responsible for different GT. The *epsF* and *epsG* genes are in the

---

**Table 2. Estimation of the growth parameters by the Baranyi model of *Streptococcus thermophilus* S-3 grown on M17 medium with lactose, glucose, or galactose as the sole carbon source**

<table>
<thead>
<tr>
<th>Carbon source</th>
<th>Maximum growth rate (h⁻¹)</th>
<th>Lag phase duration (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1% Lactose</td>
<td>0.183</td>
<td>2.926</td>
</tr>
<tr>
<td>2% Lactose</td>
<td>0.227</td>
<td>3.139</td>
</tr>
<tr>
<td>2% Glucose</td>
<td>0.041</td>
<td>5.435</td>
</tr>
<tr>
<td>4% Glucose</td>
<td>0.040</td>
<td>9.440</td>
</tr>
<tr>
<td>2% Galactose</td>
<td>0.064</td>
<td>8.292</td>
</tr>
<tr>
<td>4% Galactose</td>
<td>0.029</td>
<td>8.320</td>
</tr>
</tbody>
</table>

---

**Figure 2.** Cell growth of *Streptococcus thermophilus* S-3 grown on M17 medium with lactose, glucose, or galactose as the sole carbon source. OD₆₀₀ = optical density at 600 nm. Error bars indicate SD of the mean of triplicate.
Figure 3. Pathways of sugar metabolism for exopolysaccharide (EPS) biosynthesis (A) and EPS gene cluster (B) in *Streptococcus thermophilus* S-3 based on Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis (http://www.genome.jp/kegg). UDP = uridine diphosphate; dTDP = deoxythymidine diphosphate; FK = fructokinase; GalE = UDP-galactose-4-epimerase; GalM = galactose mutarotase; GalK = galactokinase; GalT = galactose-1-phosphate uridylyltransferase; GalU = UDP-glucose pyrophosphorylase; GIK = glucokinase; GLMU = N-acetylglucosamine-1-phosphate uridylyltransferase; GPD = glucosamine-6-phosphate deaminase; GPI = glucose-6-phosphate isomerase; GPN = glucosamine-1-phosphate N-acetyltransferase; LacZ = β-glalactosidase; nucleotide diphosphate sugars, nucleotide sugars; PEP = phosphoenolpyruvate; PGM = α-phosphoglucomutase; PLMM = phosphoglucosamine mutase; PM = phosphomutase; PTS = sugar phosphotransferase system; RmlA = dTDP-glucose pyrophosphorylase; RmlB = dTDP-glucose 4,6-dehydratase; RmlC = dTDP-4-dehydrorhamnose 3,5-epimerase; RmlD = dTDP-4-dehydrorhamnose reductase; UDP-GalNAc = UDP-N-acetylgalactosamine; UDP-GlcNAc = UDP-N-acetylglucosamine. Double-headed arrow indicates a bidirectional reaction.
GT-1 family, and epsI is in the GT-2 family. Based on sequence similarity analysis, epsF was assigned as encoding galactosyltransferase, and epsG might encode N-acetylglucosamine transferase with a low homology. The epsI gene was not identified as the specific GT. The epsH gene did not encode GT and was classified as encoding acetyltransferase (Stingele et al., 1996). Generally the substrate specificity of GT in the eps gene cluster provides valuable information on the monosaccharide composition of an EPS (Li et al., 2018). However, it is still challenging to accurately predict GT substrate specificity by bioinformatics analysis. The monosaccharide composition of EPS in S-3 was previously analyzed by our group and found to consist of N-acetylgalactosamine, galactose, and glucose in a molar ratio of 1:2:1 (Xu et al., 2018). Accordingly, we hypothesized that epsG and epsI might encode N-acetylgalactosamine transferase and glucosyltransferase, respectively, in S-3. Both deoD and orf14.9 were also present near the eps gene cluster, as reported in other strains of S. thermophilus (Tyvaert et al., 2006). Upstream of the eps gene cluster was deoD, encoding purine nucleotide phosphorylase, which is involved in nucleotide biosynthesis and catabolism. The orf14.9 gene, which is associated with the cell growth of S. thermophilus (Tyvaert et al., 2006), was downstream of the eps gene cluster.

**Effects of Different Sugars on EPS Biosynthesis in S. thermophilus S-3**

*Streptococcus thermophilus* can use a variety of monosaccharides and disaccharides as carbon and energy sources. Sugars such as lactose and glucose may be the primary materials for EPS synthesis in *S. thermophilus*. To our knowledge, few studies have reported on the systematic effects of sugars on EPS biosynthesis in *S. thermophilus*. Based on the monosaccharide composition of EPS in S-3 (Xu et al., 2018) and its nucleotide sugar metabolism (Figure 3A), lactose, glucose, and galactose were chosen to study the effects of sugars on EPS biosynthesis. Strain S-3 was cultivated in M17 medium containing these sugars. It was evident that cell growth on lactose was close to that on glucose and higher than that on galactose (Figure 2). As shown in Figure 4, EPS titers in the lactose medium and galactose medium were 228.9 ± 3.2 and 220.3 ± 3.2 mg/L, respectively, significantly (P < 0.05) higher than that in the glucose medium (108.7 ± 2.8 mg/L). These results suggest that strain S-3 prefers lactose to glucose and galactose for cell growth and EPS biosynthesis.

To understand how sugars affect EPS biosynthesis in *S. thermophilus* S-3, RNA samples isolated from S-3 grown in M17 medium with lactose or glucose were evaluated to illustrate the gene expression associated with EPS biosynthesis. The transcriptional levels for key EPS genes (epsA, epsC, epsE, epsH, and epsJ) and those for the biosynthesis of nucleotide sugars (UDP-galactose and UDP-glucose, EPS precursors; glk, pgm, galU, lacZ, lacS, galM, galE, galT, and galK) were investigated by real-time quantitative PCR. At the transcriptional level, epsA, epsC, epsE, epsH, and epsJ exhibited significantly (P < 0.05) higher (>2-fold) transcription in S-3 grown in the lactose medium than in the glucose medium (Figure 5A). This finding suggests that overexpression of eps genes resulted in the increase of the EPS titer in the lactose medium in comparison with those in the glucose medium. It confirms that lactose is the preferred sugar for EPS biosynthesis in S-3. Compared with the glucose medium, the expression of key genes responsible for UDP-glucose biosynthesis (glk, pgm, and galU) was upregulated in the lactose medium. In particular, pgm expression was obviously boosted. Interestingly, the expression of lacZ and lacS (responsible for lactose metabolism) and galM, galE, galT, and galK (responsible for UDP-galactose biosynthesis) was enhanced by more than 27-fold in the lactose medium compared with the glucose medium (Figure 5B). This implies that UDP-galactose synthesis was significantly promoted in the lactose medium in comparison with that in the glucose medium. A similar mechanism of improved EPS production was revealed by Wu and Shah (2018) using comparative mRNA-Seq analysis of *S. thermophilus* ASCC 1275 during optimized milk fermentation. In addition, Padmanabhan et al. (2018) provided transcriptomic insights into the growth phase- and sugar-associated changes in the EPS production of strain ASCC 1275, revealing that the genes responsible for the formation of UDP-glucose

![Figure 4. Influence of different carbon sources on exopolysaccharide (EPS) biosynthesis in *Streptococcus thermophilus* S-3. The data were statistically analyzed by a 1-way ANOVA using Minitab software (version 20, SPSS Inc., Chicago, IL). Error bars indicate SD of the mean of triplicates. **Asterisks indicate a statistically significant group (P ≤ 0.01).](image)
and UDP-galactose were highly upregulated when EPS production was high.

Production of EPS by LAB relates to culture conditions such as temperature, pH, nitrogen sources, and carbon sources. Among various nutrients in the culture medium, the carbon source is an important factor in the EPS yield. The preference of S-3 for lactose, rather than glucose or galactose, for EPS production found in this study is different from that reported for some Lactobacillus strains. For example, Lactobacillus delbrueckii ssp. bulgaricus NCFB 2772 showed a preference for glucose as a carbon source for producing EPS, resulting in an EPS yield 3 times that obtained from fructose (Grobben et al., 1996, 1997). Lactobacillus casei CG11 showed a higher EPS yield when grown in a medium with glucose than when grown with other sugars (Cerning et al., 1994). Production of EPS also relates to strain engineering such as engineered pathways of nucleotide sugar synthesis. For example, Levander et al. (2002) reported that overexpression of galU and pgm, key genes involved in EPS biosynthesis, increased EPS yield from 0.17 to 0.31 g/mol of carbon from lactose in S. thermophilus. Overall, the results of this study suggest that the mechanism for improving EPS biosynthesis from lactose can be attributed to the cooperation of genes involved in biosynthesis of nucleotide sugars and EPS.

CONCLUSIONS

The genes involved in carbohydrate utilization, nucleotide sugar synthesis, and EPS biosynthesis were mined by genome sequencing of S. thermophilus S-3. Genomic analysis showed that S-3 can use sucrose, mannose, glucose, galactose, and lactose. Based on phenotype analysis, S-3 prefers to ferment lactose to glucose or galactose. Compared with glucose as a carbon source, a high yield of EPS from S-3 cultivated with lactose was correlated with high transcriptional levels of the genes associated with nucleotide sugar metabolism and EPS biosynthesis. The results of this study provide a better understanding of EPS biosynthesis in S. thermophilus at the gene level. Engineering the EPS gene cluster and nucleotide sugar biosynthesis could be an efficient strategy for increasing EPS production in S. thermophilus S-3.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grant no. 31871776 and 31771956; Beijing), National Key R&D Program of China (grant no. 2018YFD0502300; Beijing), Natural Science Foundation of Shanghai (grant No. 18ZR1426800, China), “Shu Guang” project supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation (no. 15SG42, China).

REFERENCES


Figure 5. The transcriptional change in the key genes of exopolysaccharide (EPS) biosynthesis (A) and nucleotide sugar metabolism (B) by Streptococcus thermophilus S-3 cultured in M17 medium with lactose in comparison with that of glucose medium using real-time quantitative PCR. Error bars indicate SD of the mean of triplicates.