



## OPEN

SUBJECT AREAS:  
BIOTECHNOLOGY  
MICROBIOLOGYReceived  
19 February 2014Accepted  
22 April 2014Published  
15 May 2014Correspondence and  
requests for materials  
should be addressed to  
N.P.S. (npshah@hku.  
hk)

# Genomic insights into high exopolysaccharide-producing dairy starter bacterium *Streptococcus thermophilus* ASCC 1275

Qinglong Wu, Hein Min Tun, Frederick Chi-Ching Leung &amp; Nagendra P. Shah

Food and Nutritional Science, School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong.

*Streptococcus thermophilus* ASCC 1275 (ST 1275), a typical dairy starter bacterium, yields the highest known amount (~1,000 mg/L) of exopolysaccharide (EPS) in milk among the species of *S. thermophilus*. The addition of this starter in milk fermentation exhibited texture modifying properties for fermented dairy foods such as yogurt and cheese in the presence of EPS as its important metabolite. In this genomic study, a novel *eps* gene cluster for EPS assembly of repeating unit has been reported. It contains two-pair *epsC-epsD* genes which are assigned to determine the chain length of EPS. This also suggests this organism produces two types of EPSs – capsular and ropy EPS, as observed in our previous studies. Additionally, ST 1275 appears to exhibit effective proteolysis system and sophisticated stress response systems to stressful conditions, and has the highest number of four separate CRISPR/Cas loci. These features may be conducive to milk adaptation of this starter and against undesirable bacteriophage infections which leads to failure of milk fermentation. Insights into the genome of ST 1275 suggest that this strain may be a model high EPS-producing dairy starter.

Conventional dairy starter bacteria including *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Lactococcus lactis* have a long history of use in the home-made and modern manufacture of fermented dairy foods, *i.e.*, yogurt and cheese<sup>1,2</sup>. These dairy starters are able to ferment milk lactose to produce lactic acid which decreases the pH to 4.5 ~ 4.7 resulting in the coagulation of milk proteins<sup>3,4</sup>. Among these important conventional starters, *S. thermophilus* is a non-pathogenic and homofermentative facultative anaerobe, which is used for the manufacture of yogurt and certain types of cheese. There has been an increasing interest in using a novel EPS-producing *S. thermophilus* for enhancing functionalities of yogurt and cheeses<sup>5-9</sup>.

Until April 2014, six strains of *S. thermophilus* have been fully sequenced and their whole-genome sequence data are released in the NCBI Genome database<sup>10-14</sup>. Comparative genome analysis of dairy *S. thermophilus* suggests that their proteolytic activity, nitrogen metabolism, sugar utilization and transporter systems play crucial roles for their adaptation to milk environments<sup>7,12,15</sup>. In addition to the “generally recognized as safe” status of dairy *S. thermophilus* through loss-of function events such as decay and loss of virulence determinants during evolution, both lateral gene transfer (LGT) and natural competence contribute to the shaping of *S. thermophilus* genome. This kind of evolution results in diverse metabolic activities and gives new functionalities to dairy foods<sup>10,16</sup>. Common features of dairy *S. thermophilus* include rapid acidification of milk, acid tolerance, bacteriocin synthesis, lactose utilization, production of formic and folic acids, innate and adaptive immunity, bacteriophage resistance, and most importantly, exopolysaccharide (EPS) biosynthesis<sup>7,15</sup>. These features are important for dairy *S. thermophilus* as starter bacterium for its applications in milk fermentation.

Extracellular polysaccharide, also known as exopolysaccharide (EPS), produced by lactic acid bacteria (LAB) including *S. thermophilus* is generally regarded as a food-grade as it is naturally produced<sup>5,8,9</sup>. EPS may be secreted into the medium as ropy EPS, or may be attached to cell surface of the microorganism in the form of capsular EPS<sup>8</sup>. EPS has been reported to improve the viscosity and texture of yogurt and some cheeses, and to prevent syneresis in yogurt<sup>4,5,8,17-21</sup>. Moreover, EPS produced by dairy LAB is able to replace chemically modified starches or milk fat in commercial yogurt, especially set-type yogurt, to give considerable rheological effects, mouthfeel, and creaminess to fermented milk products<sup>5,20,21</sup>. Certain EPSs have also been reported to have some important



probiotic characteristics such as immunostimulative properties, anti-oxidative effects, and anti-microbial activities against pathogens<sup>22–24</sup>.

In general, EPS yield among majority of *S. thermophilus* strains varies from 20 mg/L to 600 mg/L in milk-based medium under optimal conditions<sup>9,25</sup>. Among all the reported data of EPS yield from the species of *S. thermophilus*, *S. thermophilus* ASCC 1275 (ST 1275) produced the highest known amount of EPS (~1,029 mg/L) in milk medium in presence of 0.5% whey protein concentrate when fermentation was carried out at pH 5.5 and 37°C for 24 h<sup>26</sup>. Moreover, ST 1275 produced both capsular and ropy EPS<sup>20,27</sup>. It has been documented that capsular EPS does not cause ropiness in milk products whereas ropy EPS contributes to the enhanced texture of milk products<sup>28</sup>. Our previous studies have shown that high amount of EPS produced from ST 1275 exhibited texture modifying properties in Mozzarella cheese and yogurt<sup>17–21</sup>. Additionally, the usage of ST 1275 for milk fermentation contributed to the development of low-fat or fat-free yogurt and Mozzarella cheese<sup>17,18,20,29–32</sup>. Thus, any efforts to increase EPS yield in milk would be of great significance for enhancing functionalities of fermented dairy foods.

EPS assembly of repeating unit is determined by *eps* gene cluster, which has been revealed in detail in certain species of LAB and has shown diverse gene structures so far<sup>33,34</sup>. Despite the release of *eps* gene clusters from six sequenced strains of *S. thermophilus*<sup>10–14</sup>, their data on EPS yield still remains unknown; this may be due to the commercial nature of these strains or low yield of EPS. Hence, our understanding of high EPS-producing *S. thermophilus* at genomic level is still limited. Based on our previous studies on high EPS yield from ST 1275 in milk, we used ST 1275 in the current study as a model dairy starter to demonstrate the mechanism of high EPS yield from the species of *S. thermophilus* at genomic level.

## Results

**Genome sequencing and assembly.** ST 1275 genome was sequenced by one shotgun run and one 8 kb-span paired-end run using a 454 Roche GS Junior System. A total of 72,487,271 bases generated from 158,162 raw shotgun reads and 56,596,072 bases from 152,819 raw paired-end reads were aligned into 65 contigs and 4 scaffolds, resulting in an average sequencing depth of ~62 fold. Draft genome was achieved by *de novo* assembly to produce a draft genome with 4 scaffolds containing 44 large contigs with an N50 Contig length of 100,486 bp long, indicating that this assembly was highly continuous. Only three gaps were found between the junctions of contigs, and were filled in by general PCR and Sanger sequencing method. This *de novo* shotgun paired-end pyrosequencing is able to provide high sequencing depth for microbial genome.

**General features of ST 1275.** The complete circular genome of ST 1275, which was a plasmid-free bacterium, was 1,845,495 bp with an average GC content of 39.06% (Fig. 1). A comparison of general features of five sequenced *S. thermophilus* strains and ST 1275 genome is shown in Table 1. As compared with other sequenced *S. thermophilus*, ST 1275 possessed the lowest numbers of 5 and 55 of rRNA operon and tRNA, respectively. Moreover, the highest number of four separate CRISPR/Cas loci was found in its genome suggesting that this organism may have better adaptive immunity against various bacteriophage infections.

The result of functional annotations of ST 1275 and other five sequenced *S. thermophilus* is shown in Fig. 2. In general, no major differences were found in regards to the number of genes in each functional group. Three highest numbers of genes in these six strains were found in the functional groups including those associated with protein, and amino acids and with carbohydrate metabolism. This indicates that above three functional groups are closely associated with adaptation of *S. thermophilus* to milk environment in regards to nutrients such as milk proteins and lactose.

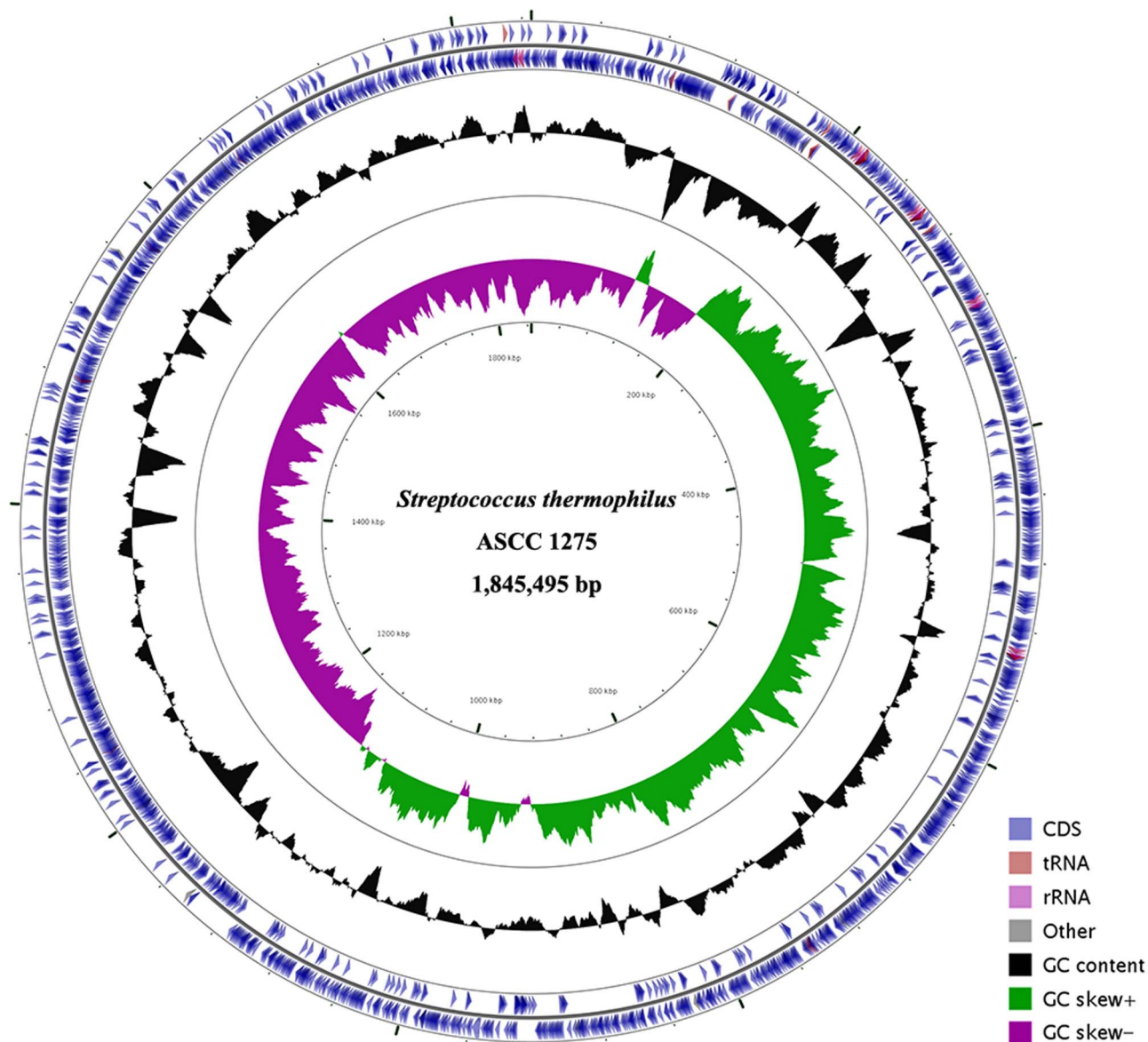
**Carbohydrate utilization and sugar transport system.** Sugar uptake, transport system and sugar hydrolases in ST 1275 are shown in Supplementary Table 1. Partial sugar metabolism involved in nucleotide sugar biosynthesis is shown in Fig. 3. Our previous studies have demonstrated that this organism was able to metabolize lactose into lactic acid efficiently resulting in rapid acidification of milk (pH 4.5–4.7) within 8 h during milk fermentation<sup>26</sup>. This is the pH at which coagulation of milk takes place, and importantly this is an acceptable fermentation period for industrial processing. In addition to utilizing lactose, galactose and glucose, ST 1275 appears to be able to ferment mannose and fructose (Supplementary Table 1 and Fig. 3). However, sucrose, mannose and fructose are the only three sugars that may be transported by specific phosphoenolpyruvate-dependent phosphotransferase systems (PEP-PTS), while lactose- and glucose-specific PEP-PTS is not available in ST 1275. Since lactose is the main sugar in milk, rapid acidification of milk by this starter is highly dependent on the utilization of lactose during milk fermentation.

Unlike limited number of hydrolases for amylose in other sequenced *S. thermophilus* strains, intact genes including one  $\alpha$ -amylase, one glucanhydrolase, three glycogen debranching proteins and two alkaline amylopullulanases were found in ST 1275 genome (Supplementary Table 1). This suggests that this organism may have an efficient amylolytic activity to break down starch<sup>35</sup>. This may be important for performing fermentation for achieving high cell-density using amylose as a cheap source of carbohydrate.

**EPS biosynthesis and comparison of *eps* gene cluster.** All essential components for EPS production including complete nucleotide sugar biosynthesis (Fig. 3) and a novel *eps* gene cluster for EPS assembly (Fig. 4) were found in ST 1275 genome. This starter contains highly conserved *epsA-epsB* which was assigned for biosynthesis regulation and *eps1C-eps1D* for determining the chain length of EPS<sup>12,36</sup>. *epsE* gene encodes a membrane-associated priming glycosyltransferase, and does not catalyze glycosidic linkage but transfers sugar-1-phosphate to undecaprenyl-phosphate-lipid carrier on the cytoplasmic face of the membrane<sup>34,37</sup>. Subsequently, *epsF*, *epsG*, *epsH*, *epsI*, *epsJ* and *epsK* encoding glycosyltransferases may transfer various nucleotide sugars including UDP-glucose, UDP-galactose, dTDP-rhamnose, UDP-GlcNAc and UDP-galactofuranose to form the repeating units in a glycosidic linkage-dependent manner<sup>34,37</sup>. Additionally, a unique UDP-galactopyranose mutase was found in this cluster for the synthesis of UDP-galactofuranose. However, chemical structure and sugar composition of repeating unit remain to be determined. Remarkably, it was for the first time that we found an additional *eps2C-eps2D* in this cluster, which may also be involved in the chain length determination. The assigned functions of polymerization and translocation of repeating units are achieved by *epsL* and *epsN*, respectively. The *epsO* and *epsP* together are possibly responsible for the phosphorylation events, while *epsQ* is assigned for the transfer of EPS between the membrane and peptidoglycan layer. It has been documented that the *orf14.9* gene distributed in all *eps* gene clusters of six strains (Fig. 4) is associated with the cell growth of *S. thermophilus*<sup>38</sup>.

In general, nucleotide sugar biosynthesis is one of the two factors for EPS yield while the *eps* gene cluster is another key factor for EPS assembly of repeating unit in lactic acid bacteria (LAB). However, various structure of *eps* gene cluster has been shown in LAB indicating that the production and chemical structure of EPS is strain-specific<sup>33,34</sup>. Interestingly, the occurrence of two-pair genes, namely *eps1C-eps1D* and *eps2C-eps2D*, for determining the chain length of EPS in ST 1275 genome implies that this starter may produce EPSs of different molecular sizes. This confirms our previous finding that ST 1275 is a producer of both capsular and ropy EPS<sup>20,27</sup>.

**Proteolytic system.** Milk is known to be a poor source of carbon and free amino acids, but contains abundance of proteins such as casein. It was found that extracellular proteinase (known as PrtS),



**Figure 1** | Circular genome map of the *S. thermophilus* ASCC 1275 chromosome. The genome of plasmid-free ST 1275 is 1,845,495 with an average GC content of 39.1%. The circular genome has been generated with the CGView Server<sup>59</sup>. The GeneBank accession number for ST 1275 genome is CP006819.

membrane transporters and intracellular peptidases contribute to the utilization of exogenous proteins by *S. thermophilus* in milk<sup>7,12</sup>. Hence, proteolysis system in ST 1275 plays a crucial role for this organism for its adaptation to milk. For extracellular proteinase, ST 1275 encodes one intact PrtS (T303\_05205), which is involved in the cleavage of casein to oligo-peptides and is only found in some strains of *S. thermophilus*. This is a key component for cell growth in milk<sup>39–41</sup>. Then, oligo-peptides and free amino acids are transported into cells by membrane amino acid/peptide transporters. Remarkably, an abundance of intracellular protease and peptidase were found in ST 1275 (Supplementary Table 2). This helps ST 1275 cells break down oligo-peptides into free amino acids for cellular metabolism or for direct utilization.

**Two-component regulatory systems.** The two-component regulatory systems (TCRSs) and related loci are shown in Supplementary Table 3. It has been documented that TCRSs are closely associated with stress and adaptive responses, bacteriocin biosynthesis, natural competence and biofilm formation<sup>42,43</sup>. Seven intact TCRSs were found in ST 1275 (Supplementary Table 3). However, certain

functions of TCRS have been poorly characterized in *S. thermophilus* and most of them have unknown functions or are involved in multiple cellular responses<sup>7</sup>.

**Stress response systems.** Acid resistance, cold and heat response, salt resistance, and oxidative stress response system for ST 1275 are shown in Supplementary Table 4. These loci presented in ST 1275 genome may play important roles for ST 1275 in adapting cells to stressful conditions, such as presence of oxygen, heat and cold, acid and salt. In addition to the TCRSs in ST 1275, additional stress regulators (T303\_00880 and T303\_09015) may be involved in the regulation of adaptive cellular responses.

Similar to other sequenced *S. thermophilus* strains, ST 1275 contains almost same number or types of heat-shock and cold-shock proteins, and oxidative stress response-related genes for bacterial fitness or performance.

For acid resistance, a proton translocating  $F_0F_1$ -ATPase system and a urease system coupled with ammonia permease were found in ST 1275 genome (Supplementary Table 4). These may contribute to internal pH homeostasis in this starter when facing extreme acidic

Table 1 | Comparison of general genome features of sequenced *S. thermophilus*

Feature	<i>S. thermophilus</i>					
	ASCC 1275	LMD-9	CNRZ1066	LMG 18311	ND03	MN-ZLW-002
Origin of strain	ASCRC (Australia)	Danisco (USA)	Yogurt isolate (France)	Yogurt isolate (UK)	Yili Group (China)	Mengniu Group (China)
Size of chromosome (bp)	1,845,495	1,856,368	1,796,226	1,796,846	1,831,949	1,848,520
No. of Plasmid	0	2	0	0	0	0
G + C content (%)	39.1	39.1	39.1	39.1	39.0	39.1
No. of ORFs (by GLIMMER v3.02)	2,253	2258	2,191	2211	2248	2258
No. of Genes	1,959	2,004	1,999	1,973	2,038	2,046
No. of CDS	1,694	1,711	1,914	1,888	1,919	1,910
Coding density (%)	77.85	79.01	90.69	88.69	88.08	87.29
No. of rRNA operons	5	6	6	6	5	5
No. of tRNAs	55	67	67	67	56	56
No. of CRISPR/Cas locus (by CRISPR finder)	4	3	1	2	3	3
GeneBank accession	CP006819	CP000419.1	CP000024.1	CP000023.1	CP002340.1	CP003499.1

environment, such as acids produced during milk fermentation. However, no loci encoding intact amino acid deiminase and decarboxylase were found in ST 1275 genome; those are also associated with maintenance of internal pH in bacteria<sup>44,45</sup>. Remarkably, urease system is only found in *S. thermophilus* among all the species of LAB, and has been found to be effective for the control of internal pH homeostasis<sup>46</sup>.

Interestingly, several salt resistance-related genes were found in ST 1275 genome. Since *S. thermophilus* is an essential starter for the manufacture of several common types of cheeses, these genes may help ST 1275 cells survive or adapt to high level of salt, especially in cheeses containing high level of salt.

**Defense system.** The loci encoding bacteriocin biosynthesis, multidrug resistance genes and competence proteins for natural transformation are shown in Supplementary Table 5. Lantibiotic is commonly produced by *S. thermophilus* as an anti-microbial weapon

against other microbes such as food-borne pathogens<sup>47</sup>. Additionally, several early and late competence genes were found in ST 1275 genome. Interestingly, it has been demonstrated that Ami (oligopeptide transporter), signal peptide and *comX* (sigma factor) are important for the induction of early competence development in *S. thermophilus*<sup>48–50</sup>. Natural competence is closely associated with LGT such as acquisition of novel genes in *S. thermophilus*<sup>2,16</sup>.

Moreover, several genes for multidrug resistance (Supplementary Table 5) including two  $\beta$ -lactamases were found in its genome. However, genes encoding above enzyme for hydrolyze  $\beta$ -lactam antibiotics is very common in LAB and recognized probiotics such as *Lactobacillus rhamnosus* GG<sup>51</sup>. The gene of  $\beta$ -lactamase may be obtained via LGT during its evolution when  $\beta$ -lactam antibiotics were common and largely used in 20th century. Other multidrug ABC transporter system may be useful for removing cytotoxic compounds. Additionally, a mucus-binding protein (T303\_03820) was found in ST 1275 genome, which indicates that this organism may

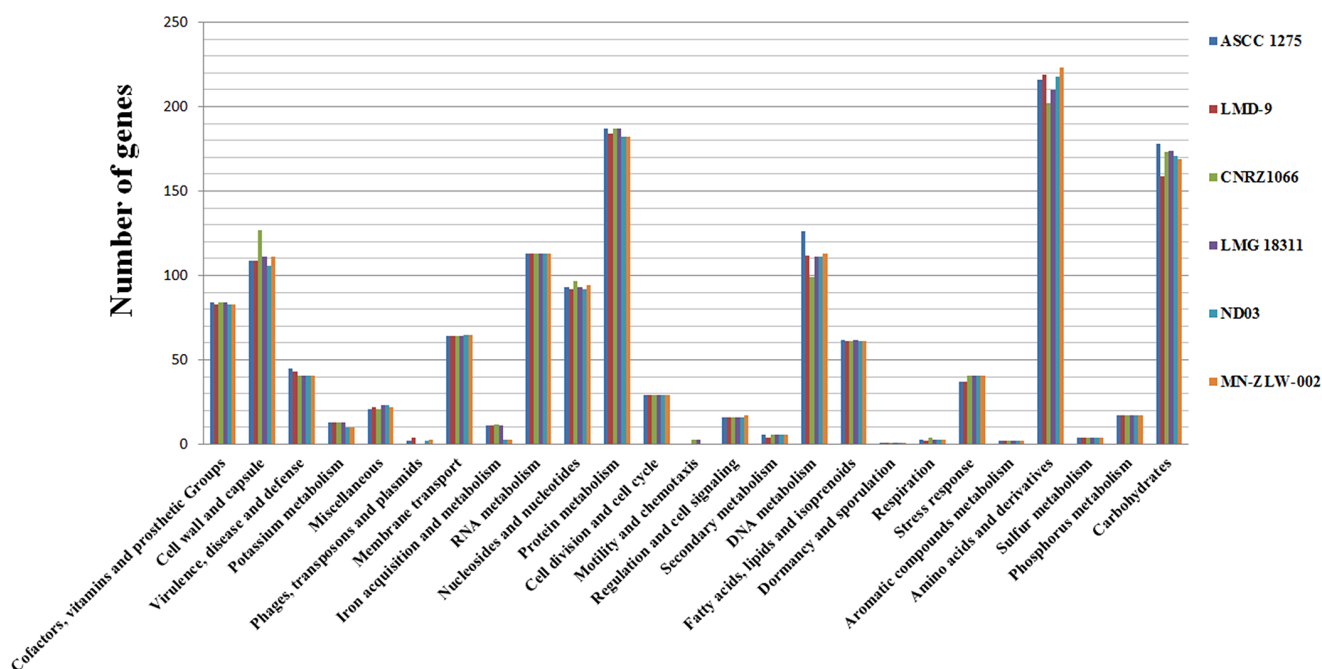
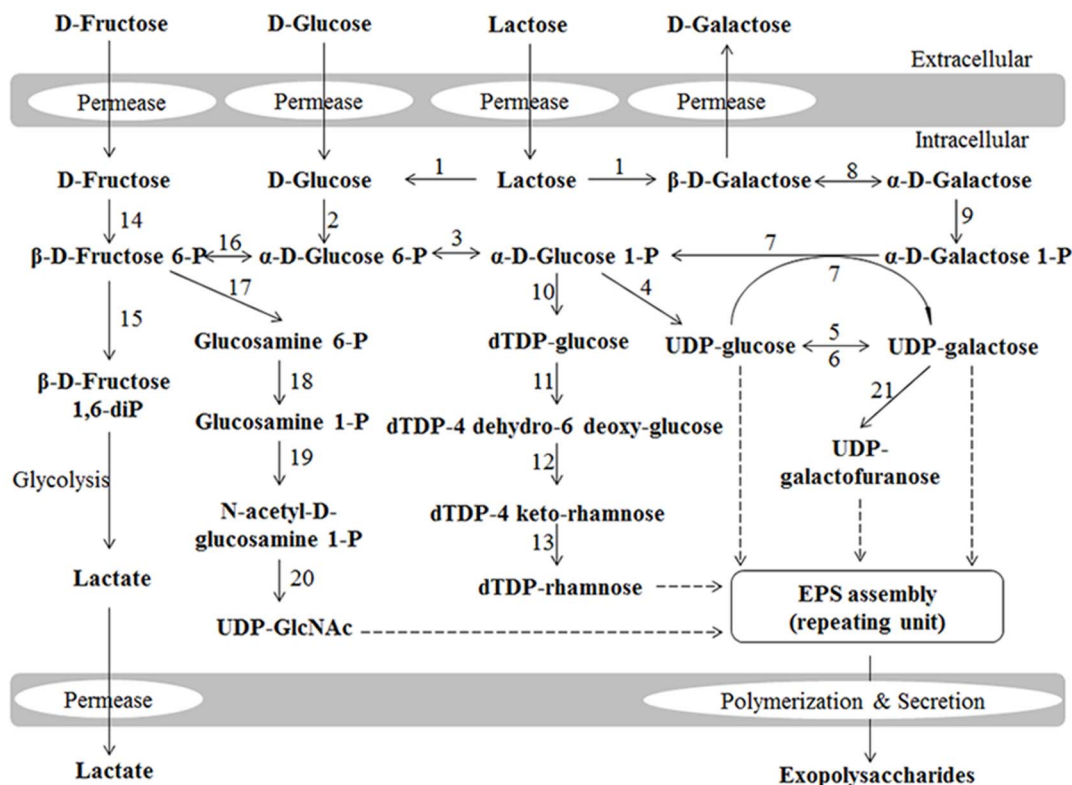


Figure 2 | Comparison of functional annotation of ST 1275 and other five sequenced *S. thermophilus* using RAST server. The nucleotide sequences of six sequenced *S. thermophilus* were uploaded into the RAST server based on SEED subsystems for functional annotations.

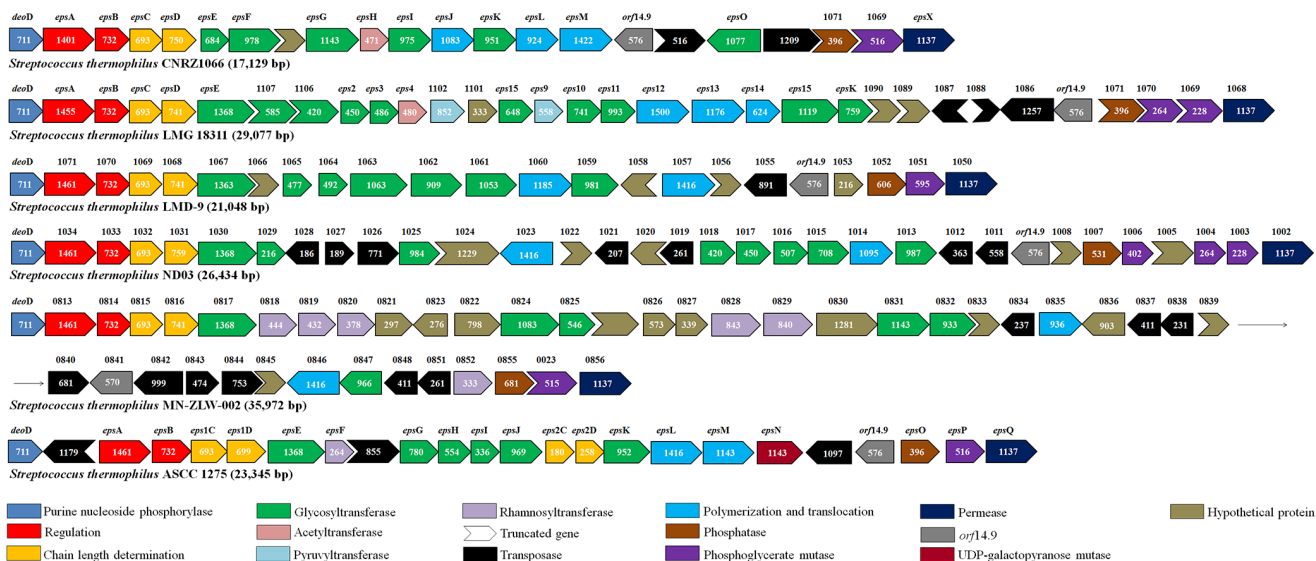


**Figure 3 | Nucleotide sugars biosynthesis for EPS production in *S. thermophilus* ASCC 1275.** The numbers refer to the enzymes involved: 1, β-Galactosidase; 2, Glucokinase; 3, Phosphoglucumutase; 4, UDP-glucose pyrophosphorylase; 5, UDP-glucose 4 epimerase; 6, UDP-galactose 4 epimerase; 7, Galactose 1-phosphate uridylyltransferase; 8, Galactose mutarotase; 9, Galactokinase; 10, dTDP-glucose pyrophosphorylase; 11, dTDP-glucose-4, 6-dehydratase; 12, dTDP-4 keto-6 deoxy-glucose 3, 5-epimerase; 13, dTDP-4 keto-L-rhamnose reductase; 14, Fructokinase; 15, 6-phosphofructokinase; 16, Phosphoglucose isomerase; 17, glutamine-fructose-6-phosphate transaminase; 18, Phosphoglucosamine mutase; 19 & 20, N-acetylglucosamine-1-phosphate uridylyltransferase (bifunctional); 21, UDP-galactopyranose mutase.

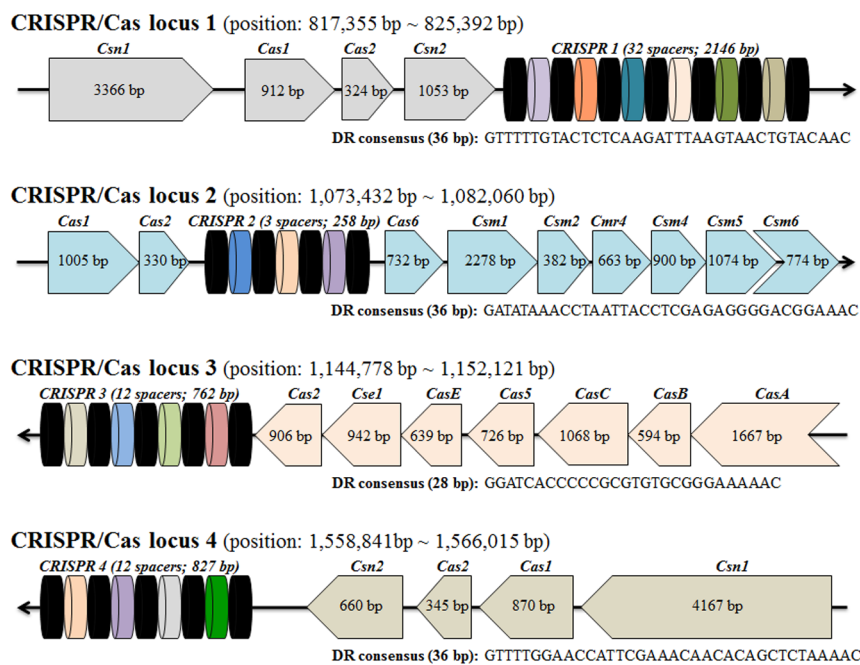
have potential as a probiotic organism to colonize and survive in the human gastrointestinal tract, especially in individuals exposed to β-lactam antibiotics.

**CRISPR/Cas system against bacteriophage infection.** Four separate CRISPR/Cas loci were predicted in the genome ST 1275 by

CRISPR finder online service (Fig. 5). Recently, CRISPR/Cas system as prokaryotic defense system against bacteriophage infections has been documented. There have been several mechanisms against bacteriophage infections in bacteria such as encounter blocks, resistance to viral absorption, penetration blocks, restriction modification and CRISPR/Cas system<sup>52</sup>. However, CRISPR/Cas system has been



**Figure 4 | Comparison of *eps* gene cluster among *S. thermophilus* ASCC 1275 and other five sequenced *S. thermophilus*.** The predicted functions of each color-coded ORF (intact or truncated) are indicated at the lower bottom panel. The size of each ORF in *eps* gene cluster is indicated in each pentagon (intact) or chevron (truncated).



**Figure 5** | Structure of CRISPR/Cas loci in *S. thermophilus* ASCC 1275. DR, direct repeat. Four different DRs were black color-coded and spacers were other color-coded. The consensus sequence and size of four DRs are indicated at lower right panel of each locus. The size of each CRISPR-associated protein in the locus is indicated in each pentagon (intact) or chevron (truncated).

widely distributed in prokaryotes as an adaptive immunity against bacteriophage infection. In addition to their innate immunity such as restriction modification system in dairy starters, adaptive immunity is very important for both dairy and starter culture industries to guard against phage infection which causes failure of milk fermentation<sup>52,53</sup>.

Interestingly, ST 1275 contains the highest numbers of CRISPR/Cas loci, possessing four CRISPR loci and 24 CRISPR-associated protein (*cas*) genes including two truncated *cas* genes, among all the sequenced strains of *S. thermophilus*. In general, three CRISPR loci are located at the downstream of *cas* genes while CRISPR2 is located in the middle of *cas* genes in CRISPR/cas locus 2. Moreover, four CRISPR loci have three different spacer numbers and four different consensus sequences of direct repeats (DRs). These diverse CRISPR/Cas loci in ST 1275 suggest that it may have a better adaptive immunity against different bacteriophages compared with those in other sequenced *S. thermophilus*. This is important for industrial manufacturing of dairy products that use this organism. In particular, CRISPR1 locus has the highest numbers of DRs and spacers when compared with other three loci. This suggests a possible effective defense mechanism to integrate novel spacers in CRISPR1 when ST 1275 is exposed to bacteriophages<sup>53</sup>. It is likely that CRISPR2 locus may have limited contribution to bacteriophage response because of less spacers. It has been demonstrated that increased expression of *cas1* and *cas2* gene was indicative of higher activity in *S. thermophilus* LMD-9 during bacteriophage response<sup>12</sup>. Thus, the distribution of *cas1* or *cas2* gene in four CRISPR/Cas loci may confer their active roles in defense system.

## Discussion

Due to the importance of EPS produced by dairy starter bacterium on the quality of fermented dairy foods, attentions have been paid to novel EPS-producing starters, especially an essential starter *Streptococcus thermophilus*<sup>1,6,7,9</sup>. Although several *S. thermophilus* genomes are available, their EPS yields are not reported, possibly due to their commercial nature or low EPS yield<sup>10–14</sup>. So far, numerous studies have been carried out for identification and characterization of *eps* gene clusters in high EPS-producing LAB while no genomic data is

available for high EPS-producing starter bacterium, especially for an important organism such as dairy *S. thermophilus*<sup>33,34</sup>. To the best of our knowledge, ST 1275 produces highest known amount of EPS (~1,000 g/L) under optimal conditions in milk as compared with other well-documented EPS-producing *S. thermophilus* strains (Supplementary Table 6), however, its regulatory mechanism for EPS yield remains poorly understood and merits further studies<sup>26</sup>. Hence, it was interesting to have insight at genomic level of ST 1275 as a model of high EPS-producing starter in the species of *S. thermophilus*.

In general, *S. thermophilus* is not able to uptake lactose via lactose-PTS but via lactose/galactose permease (*LacS*). Then lactose is hydrolyzed into glucose and galactose by  $\beta$ -D-galactosidase, and galactose is excreted into the extracellular medium by *LacS* resulting in high concentration of residual galactose in milk after fermentation<sup>6</sup>. Hence, less galactose is utilized by *S. thermophilus* as galactose is mainly metabolized for synthesis of nucleotide sugars for EPS production. However, the EPS yield in *S. thermophilus* strains is very limited. Residual galactose in cheeses such as Mozzarella cheese leads to browning during baking process of pizza made with such cheeses<sup>54</sup>. Thus, high EPS-producing *S. thermophilus* could be an ideal choice to reduce residual galactose in milk, and as well as for improving texture of dairy foods.

Since *eps* gene cluster for EPS assembly of repeating unit and nucleotide sugar biosynthesis are the two factors that have direct influence on EPS yield. Hence, we have paid attention to both of them in ST 1275 genome. Interestingly, the occurrence of *eps1C-eps1D* and *eps2C-eps2D* (Fig. 4) assigned for chain length determination indicates that this organism may assemble two types of EPSs of different molecular size. Based on our previous study that ST 1275 is a mixed producer of both capsular and ropy EPS<sup>20,27</sup>, we conclude that ST 1275 produces at least two types of EPSs. However, further work to determine the chemical structure of EPSs from ST 1275 would be important. Additionally, previous studies have demonstrated that increased gene expressions involved in nucleotide sugar biosynthesis improved the EPS production from LAB including *S. thermophilus*<sup>9,54</sup>. However, very limited information is available for the gene expression in *eps* gene cluster for EPS assembly. Since it is



common that there is only one pair *epsC-epsD* gene for chain length determination in LAB, the occurrence of two pair genes of *epsC-epsD* indicates a complex regulation of EPS production in ST 1275. In general, EPS production from LAB is cell growth-associated. However, our previous study found that optimization of cultivation conditions such as pH, temperature and addition of whey protein concentrate has resulted in a large increase in EPS yield, while no effect was observed on the cell growth of ST 1275<sup>26</sup>. This implies that gene expressions of nucleotide sugars biosynthesis or EPS assembly were possibly changed in ST 1275 under optimal conditions. Thus, mechanistic study on the regulation of EPS yield from ST 1275 merits further investigation.

Comparisons of common features including carbohydrate utilization, proteolytic system, stress response system, and defense system among the sequenced *S. thermophilus* strains suggest that ST 1275 may serve as a model for high EPS-producing dairy starter bacterium. Specifically, this strain may possess effective proteolytic system, which contributes to adaptation of this organism to milk and rapid acidification of milk. Acid resistance using unique urease system in ST 1275 may improve cell viability in extreme acidic conditions such as in yogurt. Four dependent CRISPR/Cas loci may be effective in controlling phage infection. Abundance of multidrug resistance genes and a mucus-binding protein in the cell surface may allow ST 1275 to serve as a probiotic candidate for survival and colonization in the gut and for improving gut homeostasis. The elucidation of ST 1275 genome makes this organism a model dairy starter bacterium for high EPS yield among the species of *S. thermophilus*.

## Methods

**Bacterial strain and culture conditions.** *S. thermophilus* ASCC 1275 (ST 1275), a typical dairy starter bacterium, was obtained from the Australian Starter Culture Research Center (ASCRC; now Dairy Innovation Australia Limited, Werribee, Victoria, Australia). This organism was stored at  $-80^{\circ}\text{C}$  in 10% (w/v) reconstituted skim milk containing 20% (v/v) glycerol and was activated by growing anaerobically in M17 agar (BD Company, NJ, USA) at  $37^{\circ}\text{C}$  for 24 h. After successful activation, a typical individual colony was inoculated in M17 broth containing 1% lactose and anaerobically incubated at  $37^{\circ}\text{C}$  for 18 h. Then, cells were harvested for genomic DNA extraction.

**Genomic DNA extraction.** Genomic DNA was extracted from ST 1275 using the CTAB/NaCl method according to the protocol from DOE Joint Genome Institute (JGI, <http://my.jgi.doe.gov/general/protocols.html>). Briefly, bacterial cultures were harvested by centrifugation, re-suspended in TE buffer containing lysozyme, SDS and Proteinase K, and incubated at  $37^{\circ}\text{C}$  for 1 h, followed by steps including addition of CTAB/NaCl (pre-warmed to  $65^{\circ}\text{C}$ ), incubation at  $65^{\circ}\text{C}$  for 10 min, and DNA purification using phenol/chloroform/isopropanol (25/24/1, v/v/v). Genomic DNA was precipitated and washed by adding isopropanol and 70% ethanol, respectively. Finally, DNA pellet was dried and resuspended in TE buffer containing 0.1 mg/mL of RNase. Then, the concentration and quality of genomic DNA were measured by Nanodrop-1000 UV/Vis spectrophotometer (NanoDrop Technologies, DE, USA).

**De novo shotgun paired-end pyrosequencing and genome assembly.** Shotgun sequencing, paired-end pyrosequencing and Sanger sequencing were carried out to generate the whole genome of ST 1275<sup>55,56</sup>. Briefly, shotgun sequencing was performed using 454 GS Junior System (Roche Diagnostics, CT, USA) using a GS FLX titanium rapid library preparation kit according to the manufacturer's instructions (Roche Diagnostics). One extra paired-end pyrosequencing run was carried out by using 8 kb-span library to produce a draft genome. The raw reads were *de novo* assembled into contigs using Newbler 2.7 (Roche Diagnostics). To complete the whole genome of ST 1275, primers were designed and gaps in the draft genome were filled by sequencing PCR products using ABI 3730 capillary sequencer.

**Gene prediction and annotation.** Gene annotation was carried out using NCBI Prokaryotic Genome Annotation Pipeline<sup>56</sup>. Coding sequence (CDS) prediction programs provided by GLIMMER v3.02 was used for gene prediction<sup>57</sup>. BLASTp was used to align the amino acid sequences against NCBI non-redundant database. Amino acid sequences encoded by predicted genes were searched against all proteins from complete microbial genomes, alignment length over 90% of its own length and over 60% match identity were chosen, and the best BLAST hit with highest alignment length percentage and match identity was assigned as the annotation of predicted gene<sup>55</sup>. Further annotation was obtained using the SEED-based automated annotation system provided by the RAST server<sup>58</sup>.

**Bioinformatic analysis.** CRISPR finder, a web online tool (<http://crispr.u-psud.fr/Server/>), was used for identifying CRISPR/Cas systems in bacteria. Ortholog assignment and metabolic pathway mapping of ST 1275 was executed for the amino acid sequences of CDSs using KEGG Automatic Annotation Server (KAAS; <http://www.genome.jp/tools/kaas/>), an online service based on bi-directional best hit (BBH) method.

- Iyer, R., Tomar, S. K., Maheswari, T. U. & Singh, R. *Streptococcus thermophilus* strains: multifunctional lactic acid bacteria. *Int. Dairy J.* **20**, 133–141 (2010).
- Liu, M. J., Siezen, R. J. & Nauta, A. In silico prediction of horizontal gene transfer events in *Lactobacillus bulgaricus* and *Streptococcus thermophilus* reveals proto-cooperation in yogurt manufacturing. *Appl. Environ. Microbiol.* **75**, 4120–4129 (2009).
- Heller, K. J. Probiotic bacteria in fermented foods: product characteristics and starter organisms. *Am. J. Clin. Nutr.* **73**, 374s–379s (2001).
- Lourens-Hattingh, A. & Viljoen, B. C. Yogurt as probiotic carrier food. *Int. Dairy J.* **11**, 1–17 (2001).
- Broadbent, J. R., McMahon, D. J., Oberg, C. J. & Welker, D. L. Use of exopolysaccharide-producing cultures to improve the functionality of low fat cheese. *Int. Dairy J.* **11**, 433–439 (2001).
- De Vuyst, L., Weckx, S., Ravyts, F., Herman, L. & Leroy, F. New insights into the exopolysaccharide production of *Streptococcus thermophilus*. *Int. Dairy J.* **21**, 586–591 (2011).
- Hols, P. *et al.* New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS Microbiol. Rev.* **29**, 435–463 (2005).
- Petersen, B. L., Dave, R. I., McMahon, D. J., Oberg, C. J. & Broadbent, J. R. Influence of capsular and rmp exopolysaccharide-producing *Streptococcus thermophilus* on Mozzarella cheese and cheese whey. *J. Dairy Sci.* **83**, 1952–1956 (2000).
- Vanigelgem, F. *et al.* Biodiversity of exopolysaccharides produced by *Streptococcus thermophilus* strains is reflected in their production and their molecular and functional characteristics. *Appl. Environ. Microbiol.* **70**, 900–912 (2004).
- Bolotin, A. *et al.* Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat. Biotechnol.* **22**, 1554–1558 (2004).
- Delorme, C. *et al.* Complete genome sequence of the pigmented *Streptococcus thermophilus* strain JIM8232. *J. Bacteriol.* **193**, 5581–5582 (2011).
- Goh, Y. J., Goin, C., O'Flaherty, S., Altermann, E. & Hutkins, R. Specialized adaptation of a lactic acid bacterium to the milk environment: the comparative genomics of *Streptococcus thermophilus* LMD-9. *Microb. Cell Fact.* **10**, S22 (2011).
- Kang, X. H. *et al.* Complete genome sequence of *Streptococcus thermophilus* strain MN-ZLW-002. *J. Bacteriol.* **194**, 4428–4429 (2012).
- Sun, Z. H. *et al.* Complete genome sequence of *Streptococcus thermophilus* strain ND03. *J. Bacteriol.* **193**, 793–794 (2011).
- Rasmussen, T. B. *et al.* *Streptococcus thermophilus* core genome: comparative genome hybridization study of 47 strains. *Appl. Environ. Microbiol.* **74**, 4703–4710 (2008).
- Dandoy, D. *et al.* The fast milk acidifying phenotype of *Streptococcus thermophilus* can be acquired by natural transformation of the genomic island encoding the cell-envelope proteinase PrtS. *Microb. Cell Fact.* **10**, S21 (2011).
- Amatayakul, T., Sherkat, F. & Shah, N. P. Syneresis in set yogurt as affected by EPS starter cultures and levels of solids. *Int. J. Dairy Technol.* **59**, 216–221 (2006).
- Amatayakul, T., Sherkat, F. & Shah, N. P. Physical characteristics of set yoghurt made with altered casein to whey protein ratios and EPS-producing starter cultures at 9 and 14% total solids. *Food Hydrocolloid.* **20**, 314–324 (2006).
- Bhaskaracharya, R. K. & Shah, N. P. Texture characteristics and microstructure of skim milk mozzarella cheeses made using exopolysaccharide or non-exopolysaccharide producing starter cultures. *Aust. J. Dairy Technol.* **55**, 132–138 (2000).
- Purwandari, U., Shah, N. P. & Vasiljevic, T. Effects of exopolysaccharide-producing strains of *Streptococcus thermophilus* on technological and rheological properties of set-type yoghurt. *Int. Dairy J.* **17**, 1344–1352 (2007).
- Zisu, B. & Shah, N. P. Low-fat mozzarella as influenced by microbial exopolysaccharides, preacidification, and whey protein concentrate. *J. Dairy Sci.* **88**, 1973–1985 (2005).
- Fanning, S. *et al.* Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2108–2113 (2012).
- Ruas-Madiedo, P., Hugenholtz, J. & Zoon, P. An overview of the functionality of exopolysaccharides produced by lactic acid bacteria. *Int. Dairy J.* **12**, 163–171 (2002).
- Welman, A. D. & Maddox, I. S. Exopolysaccharides from lactic acid bacteria: perspectives and challenges. *Trends Biotechnol.* **21**, 269–274 (2003).
- De Vuyst, L. *et al.* Exopolysaccharide-producing *Streptococcus thermophilus* strains as functional starter cultures in the production of fermented milks. *Int. Dairy J.* **13**, 707–717 (2003).
- Zisu, B. & Shah, N. P. Effects of pH, temperature, supplementation with whey protein concentrate, and adjunct cultures on the production of exopolysaccharides by *Streptococcus thermophilus* 1275. *J. Dairy Sci.* **86**, 3405–3415 (2003).



27. Zisu, B. *Impact of pre-acidification, fat replacers and exopolysaccharide producing starter cultures on functionality of low fat Mozzarella cheese*. PhD thesis, Victoria University (2005).
28. Hassan, A. N. ADISA Foundation Scholar Award: Possibilities and challenges of exopolysaccharide-producing lactic cultures in dairy foods. *J. Dairy Sci.* **91**, 1282–1298 (2008).
29. Amatayakul, T., Halmos, A. L., Sherkat, F. & Shah, N. P. Physical characteristics of yoghurts made using exopolysaccharide-producing starter cultures and varying casein to whey protein ratios. *Int. Dairy J.* **16**, 40–51 (2006).
30. Ramchandran, L. & Shah, N. P. Effect of exopolysaccharides on the proteolytic and angiotensin-I converting enzyme-inhibitory activities and textural and rheological properties of low-fat yogurt during refrigerated storage. *J. Dairy Sci.* **92**, 895–906 (2009).
31. Ramchandran, L. & Shah, N. P. Characterization of functional, biochemical and textural properties of synbiotic low-fat yogurts during refrigerated storage. *LWT-Food Sci. Technol.* **43**, 819–827 (2010).
32. Zisu, B. & Shah, N. P. Texture characteristics and pizza bake properties of low-fat Mozzarella cheese as influenced by pre-acidification with citric acid and use of encapsulated and ropy exopolysaccharide producing cultures. *Int. Dairy J.* **17**, 985–997 (2007).
33. Broadbent, J. R., McMahon, D. J., Welker, D. L., Oberg, C. J. & Moineau, S. Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. *J. Dairy Sci.* **86**, 407–423 (2003).
34. Jolly, L. & Stingle, F. Molecular organization and functionality of exopolysaccharide gene clusters in lactic acid bacteria. *Int. Dairy J.* **11**, 733–745 (2001).
35. Slos, P., Bourquin, J. C., Lemoine, Y. & Mercenier, A. Isolation and characterization of chromosomal promoters of *Streptococcus salivarius* subsp. *thermophilus*. *Appl. Environ. Microbiol.* **57**, 1333–1339 (1991).
36. Minic, Z. *et al.* Control of EpsE, the phosphoglycosyltransferase initiating exopolysaccharide synthesis in *Streptococcus thermophilus*, by EpsD tyrosine kinase. *J. Bacteriol.* **189**, 1351–1357 (2007).
37. Lebeer, S. *et al.* Identification of a gene cluster for the biosynthesis of a long, galactose-rich exopolysaccharide in *Lactobacillus rhamnosus* GG and functional analysis of the priming glycosyltransferase. *Appl. Environ. Microbiol.* **75**, 3554–3563 (2009).
38. Tyvaert, G., Morel, C., Joly, J. P., Decaris, B. & Charron-Bourgoin, F. The constant gene orf14.9, which belongs to the variable eps (exopolysaccharide) cluster, is involved in the cell growth of *Streptococcus thermophilus*. *Can. J. Microbiol.* **52**, 908–912 (2006).
39. Courtin, P., Monnet, V. & Rul, F. Cell-wall proteinases PrtS and PrtB have a different role in *Streptococcus thermophilus*/*Lactobacillus bulgaricus* mixed cultures in milk. *Microbiology* **148**, 3413–3421 (2002).
40. Fernandez-Espla, M. D., Garault, P., Monnet, V. & Rul, F. *Streptococcus thermophilus* cell wall-anchored proteinase: release, purification, and biochemical and genetic characterization. *Appl. Environ. Microbiol.* **66**, 4772–4778 (2000).
41. Shahbal, S., Hemme, D. & Renault, P. Characterization of a Cell Envelope-Associated Proteinase Activity from *Streptococcus thermophilus* H-strains. *Appl. Environ. Microbiol.* **59**, 177–182 (1993).
42. Hoch, J. A. Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **3**, 165–170 (2000).
43. Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-component signal transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000).
44. De Biase, D. & Pennacchietti, E. Glutamate decarboxylase-dependent acid resistance in orally acquired bacteria: function, distribution and biomedical implications of the gadBC operon. *Mol. Microbiol.* **86**, 770–786 (2012).
45. Lu, P. L. *et al.* L-glutamine provides acid resistance for *Escherichia coli* through enzymatic release of ammonia. *Cell Res.* **23**, 635–644 (2013).
46. Zotta, T., Ricciardi, A., Rossano, R. & Parente, E. Urease production by *Streptococcus thermophilus*. *Food Microbiol.* **25**, 113–119 (2008).
47. Siezen, R. J., Kuipers, O. P. & deVos, W. M. Comparison of lantibiotic gene clusters and encoded proteins. *Anton. Leeuw.* **69**, 171–184 (1996).
48. Gardan, R., Besset, C., Guillot, A., Gitton, C. & Monnet, V. The oligopeptide transport system is essential for the development of natural competence in *Streptococcus thermophilus* strain LMD-9. *J. Bacteriol.* **191**, 4647–4655 (2009).
49. Fontaine, L. *et al.* A novel pheromone quorum-sensing system controls the development of natural competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. *J. Bacteriol.* **192**, 1444–1454 (2010).
50. Gardan, R. *et al.* Extracellular life cycle of ComS, the competence-stimulating peptide of *Streptococcus thermophilus*. *J. Bacteriol.* **195**, 1845–1855 (2013).
51. Mathur, S. & Singh, R. Antibiotic resistance in food lactic acid bacteria - a review. *Int. J. Food Microbiol.* **105**, 281–295 (2005).
52. Abendon, S. T. Bacterial ‘immunity’ against bacteriophages. *Bacteriophage* **2**, 50–54 (2012).
53. Mills, S. *et al.* CRISPR analysis of bacteriophage-insensitive mutants (BIMs) of industrial *Streptococcus thermophilus*- implications for starter design. *J. Appl. Microbiol.* **108**, 945–955 (2010).
54. Neves, A. R. *et al.* Towards Enhanced Galactose Utilization by *Lactococcus lactis*. *Appl. Environ. Microbiol.* **76**, 7048–7060 (2010).
55. Li, Y. F. *et al.* Complete genome analysis of a *Haemophilus parasuis* serovar 12 strain from China. *PLOS ONE* **8**, e68350 (2013).
56. Wong, C. F. *et al.* Genome sequence of *Pseudomonas mendocina* DLHK, isolated from a biotrickling reactor. *J. Bacteriol.* **194**, 6326–6326 (2012).
57. Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**, 4636–4641 (1999).
58. Aziz, R. K. *et al.* The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
59. Grant, J. R. & Stothard, P. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.* **36**, W181–W184 (2008).

## Acknowledgments

We would like to thank the Australian Starter Culture Research Center (now Dairy Innovation Australia Limited, Werribee, Victoria, Australia) for providing *Streptococcus thermophilus* ASCC 1275.

## Author contributions

N.P.S. initiated the genome project of *Streptococcus thermophilus* ASCC 1275; N.P.S. and Q.W. designed and coordinated this genome project; H.M.T. provided valuable suggestions for this project and for manuscript preparation; Q.W. prepared genomic DNA for shotgun and paired-end sequencing, filled in the gaps of draft genome, performed bioinformatic analysis of the genome, analyzed metabolic pathways, prepared all figures and tables, and wrote the draft manuscript; F.C.L. coordinated genome sequencing and assembly of ST 1275 in his lab; N.P.S. reviewed, revised, and edited the manuscript and provided valuable suggestions.

## Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wu, Q., Tun, H.M., Leung, F.C.-C. & Shah, N.P. Genomic insights into high exopolysaccharide-producing dairy starter bacterium *Streptococcus thermophilus* ASCC 1275. *Sci. Rep.* **4**, 4974; DOI:10.1038/srep04974 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. The images in this article are included in the article's Creative Commons license, unless indicated otherwise in the image credit; if the image is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the image. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>