Food Suppl Biomater Health. 2021 Sep;1(3):e28
https://doi.org/10.52361/fsbh.2021.1.e28
pISSN 2765-4362·eISSN 2765-4699

FSBH Food Supplements and Biomaterials for Health

Original Article

Check for updates

# Complete Genome Sequencing and Comparative Genomic Analysis of *Streptococcus thermophilus* CKDB027, a Promising Probiotic Bacterial Strain

**Myung-Soo Kim**,[1] **Bonggyu Min**,[1] **Insu Kim**,[1] **Yoo-Jin Kwon**,[2] **Sung-Yurb Park**,[2] **Byung-Yong Kim**,[1] **Byoung-Kook Kim**,[2] **Sung Sun Park** [1]

[1]R&D Center, Chong Kun Dang Healthcare, Seoul, Korea
[2]Probiotics Research Laboratory, Chong Kun Dang Bio Research Institute, Ansan, Korea

OPEN ACCESS

**Received:** Aug 16, 2021
**Revised:** Sep 19, 2021
**Accepted:** Sep 22, 2021

**Correspondence:**
**Sung Sun Park, PhD**
R&D Center, Chong Kun Dang Healthcare, 5–6th FL., Chong Kun Dang Healthcare Bldg., 383 Cheongpa-ro, Yongsan-gu, Seoul 07249, Korea.
E-mail: saintsun@ckdhc.com



**ORCID iDs**
Myung-Soo Kim
https://orcid.org/0000-0002-2789-4599
Bonggyu Min
https://orcid.org/0000-0001-6868-3635
Insu Kim
https://orcid.org/0000-0002-6282-4457
Sung-Yurb Park
https://orcid.org/0000-0002-8080-576X
Byung-Yong Kim
https://orcid.org/0000-0002-4229-8859
Byoung-Kook Kim
https://orcid.org/0000-0002-6783-526X
Sung Sun Park
https://orcid.org/0000-0003-4163-8927

## ABSTRACT

*Streptococcus thermophilus* is a bacterial species widely used in the food industry for producing dairy fermented foods as well as probiotics. In this study, the whole genome of the *S. thermophilus* CKDB027 strain used in probiotic products was sequenced, and the genetic features related to its safety and functionality were determined. The CKDB027 strain has a single circular chromosome of approximately 1.88 Mb with 39.0% GC content. The genome study has shown the absence of antibiotic resistance and virulence related genes. To identify the unique genetic features of CKDB027, genomic comparisons with other 18 *S. thermophilus* used in probiotic industry were conducted. The results showed that the CKDB027 strain shared a common ancestor with several strains isolated from milk or yogurt. The CKDB027 strain possesses genes related to lactose catabolism, proteolysis, stress resistance, defense system, and adherence that are major function of probiotics. In addition, a gene cluster producing exopolysaccharides was detected. These findings indicate that the *S. thermophilus* CKDB027 strain is genetically safe and has beneficial genetic features for human health to use in dairy industry and health functional foods.

**Keywords:** Probiotics; *Streptococcus thermophilus*; Whole genome sequencing; Comparative genomic analysis; Safety assessment

## INTRODUCTION

Probiotics are live microorganisms with beneficial effects on human health, such as enhancing digestive functions, increasing the proportion of beneficial intestinal bacteria, and inhibiting harmful bacterial growth.[1] In addition, they have various effects on human diseases, including improvement in immune diseases, such as atopic dermatitis and asthma, and reduction in blood cholesterol and body fat.[2] In line with the reports of such efficacy, the probiotics industry has shown rapid annual growth and active research and development of probiotics with new functionality. Recently, numerous studies have also applied probiotics to the prevention of coronavirus disease 2019 disease.[3]

**Disclosure**
The authors have no potential conflicts of interest to disclose.

**Author Contributions**
Conceptualization: Park SS; Data curation: Kim I; Formal analysis: Kim MS, Min B; Project administration: Park SS; Resources: Kim BK, Park SY, Kim I; Writing - original draft: Kim MS, Min B; Writing - review & editing: Kim BY, Kwon YJ.

Nevertheless, problems related to safety and side effects have been pointed out after the consumption of probiotics. For example, certain strains of *Enterococcus faecium* may exhibit antibiotic resistance (AR) or cause opportunistic infections; therefore, they are not prescribed to patients or older adults with low immunity.[4] The AR gene (ARG) may be transmitted to a pathogenic species in the body through mobile genetic elements, such as a bacteriophage, plasmid, or transposon.[5] Therefore, the presence of virulence genes or ARGs should be rigorously tested via genomic analysis. The US Food and Drug Administration (FDA) and the European Food Safety Authority (EFSA) apply the evaluation system of Generally Recognized as Safe (GRAS) and Qualified Presumption of Safety (QPS) for the safety management of microorganisms used in probiotics.

In the screening of a novel probiotic strains, the use of strains with ARGs or virulence genes in probiotics is prohibited by the EFSA and hence excluded. Thus, the analysis of the genetic characteristics related to virulence and AR is an essential process in the selection of beneficial strains. In addition, through whole-genome sequencing, the bioactivity and health benefits can be predicted at the genetic level. For example, the possibility of producing exopolysaccharides (EPSs) that contribute to enhancing food texture as well as various bioactive functions including lowering cholesterol, anti-oxidation, anti-virus, and enhancing immune system can be determined in advance.[6]

*Streptococcus thermophilus* is a species of lactic acid bacteria, most frequently used as a starter in fermented dairy products, such as yogurt and cheese. *S. thermophilus* rapidly converts lactose to lactic acid to create the refreshing acidic flavor while preventing the growth of acid-susceptible pathogens and harmful microorganisms.[7] *S. thermophilus* has a high functional resemblance to *Lactococcus lactis* used in probiotics, but it is phylogenetically far closer to the species *Streptococcus pneumoniae*, *Streptococcus pyogenes*, and *Streptococcus agalactiae*, which are harmful to human health.[8] However, with the accumulation of numerous studies proving the safety of the species, *S. thermophilus* is accepted as satisfying the GRAS and QPS status by the FDA and EFSA, respectively.[9]

Through whole-genome analysis, the safety and functionality of *S. thermophilus* have been evaluated. Studies comparing the genome of *S. thermophilus* against the streptococcal pathogens have shown that pathogenic determinants are absent or inactivated as pseudogenes.[10] ARGs were not detected in most strains. Although recent studies have reported certain strains with resistance, analysis of their genetic characteristics showed that the possibility of transmission is extremely low.[11]

In this study, the whole genome of the *S. thermophilus* CKDB027 strain used in commercial probiotic products was sequenced to analyze the general genomic features of the strain. Safety was verified by genome annotation to identify the presence of AR or virulence, and the genes related to functionality were identified. Furthermore, comparing with various *S. thermophilus* used in probiotic products revealed phylogenetic similarities and functional differences. The whole-genome sequencing results showed that the CKDB027 strain is genetically safe and has several beneficial functional genes. The results indicated that the *S. thermophilus* CKDB027 strain can be used in the food industry as probiotic products with verified safety.

# METHODS

## Strain isolation, growth, and identification

The CKDB027 strain was isolated from raw milk by applying the methods described in a previous study.[12] After isolation from the sample, the bacterial cells were cultured in sterile deMan Rogosa Sharpe (Becton Dickinson Co., Sparks, MD, USA) at 37°C for 24 hours. For accurate identification of the isolated strain, the 16S rRNA gene sequence was analyzed. The preparation of genomic DNA, polymerase chain reaction (PCR) amplification of 16S rRNA gene sequence, and sequencing were carried out following Macrogen Identification Service (Macrogen Inc., Seoul, Korea). The genomic DNA for PCR analysis was prepared using a DNA extraction kit (Bioneer Corp., Daejeon, Korea). For the PCR amplification of the full length (V1–V9) 16S rRNA gene sequence, the universal bacterial primer sets 27F 5′ (AGA GTT TGA TCM TGG CTC AG) 3′ and 1492R 5′ (TAC GGY TAC CTT GTT ACG ACT T) 3′ were used and sequencing was performed by ABI PRISM 3730XL DNA Analyzer and BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific Inc., Waltham, MA, USA). The strain identification was confirmed by total regions (V1–V9) of 16S rRNA gene sequence using EzBioCloud (https://www.ezbiocloud.net/) database.[13]

## Genome sequencing and assembly

The whole genome of the CKDB027 strain was sequenced using the Pacific Biosciences (PacBio) RS II (Pacific Biosciences Inc., Menlo Park, CA, USA) and Illumina iSeq 100 system (Illumina, Inc., San Diego, CA, USA). The data collected from the PacBio RS II instrument was assembled based on the Canu v1.6 protocol.[14] In addition, for the completion of more accurate whole genome sequencing, Illumina iSeq 100 was used in the resequencing of the CKDB027 genomic data. After mapping 150 base pairs (bps) paired-end reads from the iSeq 100 system based on the BWA-MEM algorithm,[15] the single base errors were corrected using the GATK haplotypecaller.[16]

## Gene annotation

The assembled genome was annotated using the Rapid Annotation using Subsystem Technology.[17] In addition, the EzBioCloud platfrom (https://www.ezbiocloud.net/) was used in the genomic mapping and the analysis of function based on the Cluster of Orthologous Groups (COG) category.[13] Using PHASTER, the prophage sequence was predicted, and CRISPRCasFinder was used to confirm the clustered regularly interspaced short palindromic repeats (CRISPR) region.[18] The genomic islands (GIs) were predicted using IslandViewer 4.[19] Comprehensive Antibiotic Resistance Database[20] was used to detect the ARGs (cut-off value: identity 70%, coverage 60%), and the virulence genes were detected based on the *Streptococcus* virulence factors using the Virulence Factor Database.[21] The predicted genes for bacteriocin and cluster for secondary metabolite biosynthesis were identified using BAGEL4.[22] The assignment of protein function for the open reading frames (ORFs) was performed manually using blastp (https://blast.ncbi.nlm.nih.gov/).[23]

## Comparative genomic analysis

Based on the National Center for Biotechnology Information (NCBI) genome database (https://www.ncbi.nlm.nih.gov/genome/), the complete genomes were selected for the 18 *S. thermophilus* strains used in probiotic products.[6,24] The average nucleotide identity (ANI) was calculated using JSpecies.[25] Based on the ANI value, the phylogenetic tree was built and visualized using MEGA-X software.[26] The Pan-genome Orthologous Groups analysis was performed using the EzBioCloud Comparative Genomics Database (https://www.ezbiocloud.net/) at a cut-off value of 90%.[13]


https://doi.org/10.52361/fsbh.2021.1.e28

# RESULTS

## General features

We obtained entire sequence of 16S rRNA gene amplicon including full fragments of V1–V9 with 1,512 bps. As a result of 16S rRNA gene sequence analysis, the CKDB027 strain showed 99.9% similarity to the type strain, *S. thermophilus* ATCC 19258$^T$. The phylogenetic tree based on the 16S rRNA gene sequence showed that CKDB027 strain is close to *S. thermophilus* species (**Supplementary Fig. 1**).

The draft genome of the CKDB027 strain was obtained by de novo assembly of reads from PacBio RS II, and Illumina iSeq 100 system. The reads from the shotgun whole genome sequencing were aligned to draft genome, and the mean coverage was 229.4× with all regions mapped. The final genome was completed by error correcting with the shotgun whole genome sequencing reads. Whole-genome sequencing showed that the CKDB027 strain has a single, circular chromosome with a genome size of 1,884,450 bps and 39.0% GC content (**Fig. 1**). Gene annotation predicted 2,126 coding sequences, which included 384 uncharacterized sequences. In addition, 74 RNA sequences, containing 59 tRNA and 15 rRNA, were identified (**Table 1**).

The COG functional categorization of a total of 1,841 ORFs revealed the following 3 major categories, excluding 439 ORFs in category S with unknown functions: 254 ORFs in category L (replication, recombination, and repair), 200 ORFs in category E (amino acid transport and metabolism), and 148 ORFs in category J (translation, ribosomal structure, and biogenesis) (**Table 2**).

**Table 1.** General genomic features of *S. thermophilus* CKDB027 strain

| Features | CKDB027 |
|---|---|
| Genome size (bps) | 1,884,450 |
| Contig number | 1 |
| GC content (%) | 39.0 |
| Total No. of genes | 2,200 |
| No. of CDS | 2,126 |
| Hypothetical protein (%) | 25.6 |
| tRNA genes | 59 |
| rRNA genes | 15 |

bp = base pair, CDS = coding sequences.

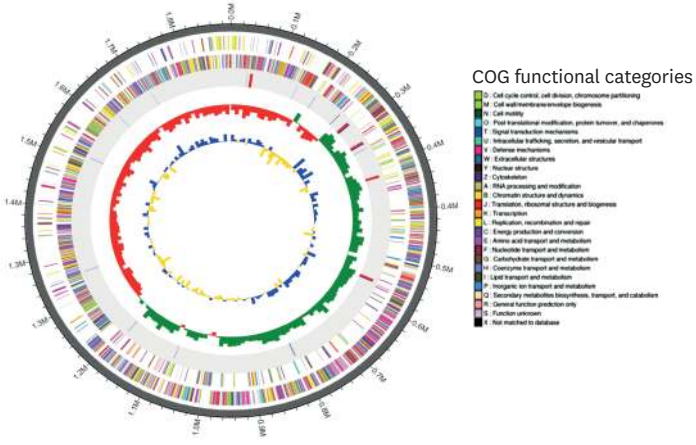


**Fig. 1.** Circular genome map of *S. thermophilus* CKDB027. From outside to inner of the circle: CDS on forward strand colored by COG category assignment, CDS on reverse strand colored by COG category assignment, RNA, GC skewness, and GC ratio.
CDS = coding sequences, COG = Cluster of Orthologous Groups.


https://doi.org/10.52361/fsbh.2021.1.e28

**Table 2.** Number of genes associated with COG functional categories of *S. thermophilus* CKDB027 strain

| COG | Function | Count | Percentage (%) |
|---|---|---|---|
| S | Function unknown | 439 | 23.8 |
| L | Replication, recombination and repair | 254 | 13.8 |
| E | Amino acid transport and metabolism | 200 | 10.9 |
| J | Translation, ribosomal structure and biogenesis | 148 | 8.0 |
| M | Cell wall/membrane/envelope biogenesis | 99 | 5.4 |
| K | Transcription | 96 | 5.2 |
| G | Carbohydrate transport and metabolism | 95 | 5.2 |
| P | Inorganic ion transport and metabolism | 87 | 4.7 |
| F | Nucleotide transport and metabolism | 74 | 4.0 |
| O | Posttranslational modification, protein turnover, chaperones | 63 | 3.4 |
| V | Defense mechanisms | 59 | 3.2 |
| H | Coenzyme transport and metabolism | 44 | 2.4 |
| C | Energy production and conversion | 42 | 2.3 |
| T | Signal transduction mechanisms | 42 | 2.3 |
| I | Lipid transport and metabolism | 32 | 1.7 |
| U | Intracellular trafficking, secretion, and vesicular transport | 29 | 1.6 |
| D | Cell cycle control, cell division, chromosome partitioning | 17 | 0.9 |
| Q | Secondary metabolites biosynthesis | 15 | 0.8 |
| N | Cell motility | 5 | 0.3 |
| Z | Cytoskeleton | 1 | 0.1 |

COG = Cluster of Orthologous Groups.

In the single chromosome of the CKDB027 strain, one complete and 3 incomplete prophage regions were identified (**Supplementary Table 1**). In one complete prophage region (region 4), 25 phage-related genes were found, which were thus encoded as the attachment sites (*attL* and *attR*), integrase, and 7 putative transposase genes of the prophage genome. From the 3 incomplete phage regions, we identified a total of 46 phage-related genes, including tail, *attL*, *attR*, and 7 putative transposase genes.

## Safety properties: ARGs, virulence factors and GIs

The complete genome of the *S. thermophilus* CKDB027 strain did not show any genes encoding the AR proteins: penicillin, glycopeptide, aminoglycoside, macrolide, lincosamides, tetracycline, and chloramphenicol. In addition, no virulence genes related to cytolysin (*cylA*, *cylB*, *cylD*, *cylE*, *cylF*, *cylG*, *cylI*, *cylJ*, *cylK*, *cylX*, and *cylZ*), aggregation substance (*asa1* and *asp1*), hyaluronidase (*hyl*), or gelatinase (*gelE*) were detected, and other virulence genes, namely, *spyA*, *aspC*, *cfa*, *cfb*, *ply*, and *slo*, were also absent.

To examine the transduction ability through horizontal gene transfer, GIs were identified, and total 25 GIs, including 11 GIs based on the SIGI-HMM algorithm and 14 GIs based on the IslandPath-DIMOB algorithm, were detected (**Supplementary Fig. 2**). However, neither virulence genes nor ARGs were detected in the GIs. These results confirmed that the *S. thermophilus* CKDB027 genome did not have ARGs or virulence genes.

## Comparison with other *S. thermophilus* strains

From the NCBI database, 18 complete genomes of *S. thermophilus* strains applied in the food industry were obtained for the analysis of the phylogenetic relationship with the CKDB027 strain. The genomic data of the 18 collected strains are summarized in **Table 3**. All 18 *S. thermophilus* strains were confirmed to exhibit ≥ 99% similarity to the CKDB027 strain based on the ANI values (**Fig. 2**). The highest similarity was observed with the LMD-9 (99.94%) and SMQ-301 (99.95%) strains, and the CKDB027 strain shared 1,742 and 1,737 genes with the 2 strains, respectively. The results also indicated 87 unique genes that were not shared with other strains (COG category; 17 category L [replication and recombination], 16 category S


https://doi.org/10.52361/fsbh.2021.1.e28

**Table 3.** General genomic features of 19 *S. thermophilus* strains

| Strain | Origin | Length (bps) | CDS | tRNA | rRNA | Hypothetical protein (%) | Assigned function (%) | GC (%) | Reference |
|---|---|---|---|---|---|---|---|---|---|
| CKDB027 | Raw milk | 1,884,450 | 2,126 | 59 | 15 | 25.6 | 74.4 | 39.0 | - |
| APC151 | Fish intestine | 1,839,134 | 2,072 | 67 | 18 | 19.0 | 81.0 | 38.3 | Linares et al.[27] |
| ASCC 1275 | Milk | 1,845,495 | 2,073 | 55 | 15 | 18.0 | 82.0 | 39.1 | Wu et al.[6] |
| ATCC 19258 | Milk | 2,102,268 | 2,385 | 56 | 15 | 23.0 | 77.0 | 39.0 | Cho et al.[28] |
| CNRZ1066 | Commercial yogurt | 1,796,226 | 2,016 | 67 | 18 | 17.9 | 82.1 | 39.1 | Bolotin et al.[29] |
| DGCC 7710 | Dairy culture | 1,851,207 | 2,072 | 56 | 15 | 18.2 | 81.8 | 39.0 | Hatmaker et al.[30] |
| JIM 8232 | Raw milk | 1,929,905 | 2,128 | 67 | 18 | 20.1 | 79.9 | 38.9 | Delrome et al.[31] |
| KLDS 3.1003 | Traditional yogurt | 1,899,956 | 2,121 | 68 | 18 | 19.5 | 80.5 | 38.9 | Evivie et al.[32] |
| KLDS SM | Traditional yogurt | 1,856,787 | 2,074 | 67 | 18 | 18.2 | 81.8 | 39.1 | Li et al.[33] |
| LMD-9 | Yogurt | 1,856,368 | 2,087 | 67 | 18 | 18.3 | 81.7 | 39.1 | Markarova et al.[34] |
| LMG 18311 | Commercial yogurt | 1,796,846 | 2,021 | 67 | 18 | 18.1 | 81.9 | 39.1 | Bolotin et al.[29] |
| MN-BM-A01 | Traditional yogurt block | 1,876,516 | 2,146 | 67 | 18 | 18.1 | 81.9 | 39.1 | Bai et al.[35] |
| MN-BM-A02 | Dairy fan | 1,850,434 | 2,074 | 57 | 15 | 18.1 | 81.9 | 39.0 | Shi et al.[36] |
| MN-ZLW-002 | Traditional yogurt block | 1,848,520 | 2,078 | 57 | 15 | 18.5 | 81.5 | 39.1 | Kang et al.[37] |
| NCTC12958 | Milk | 2,102,271 | 2,388 | 56 | 15 | 22.9 | 77.1 | 39.0 | Cho et al.[28] |
| ND03 | Naturally fermented yak milk | 1,831,949 | 2,065 | 57 | 15 | 19.0 | 81.0 | 39.0 | Sun et al.[38] |
| SMQ-301 | Milk | 1,861,792 | 2,076 | 67 | 18 | 18.1 | 81.9 | 39.1 | Labrie et al.[39] |
| ST106 | Milk | 1,856,083 | 2,165 | 67 | 18 | 17.8 | 82.2 | 39.3 | Renye et al.[40] |
| ST109 | Milk | 1,788,866 | 2,016 | 67 | 18 | 18.0 | 82.0 | 39.2 | Renye et al.[40] |

bp = base pair, CDS = coding sequences

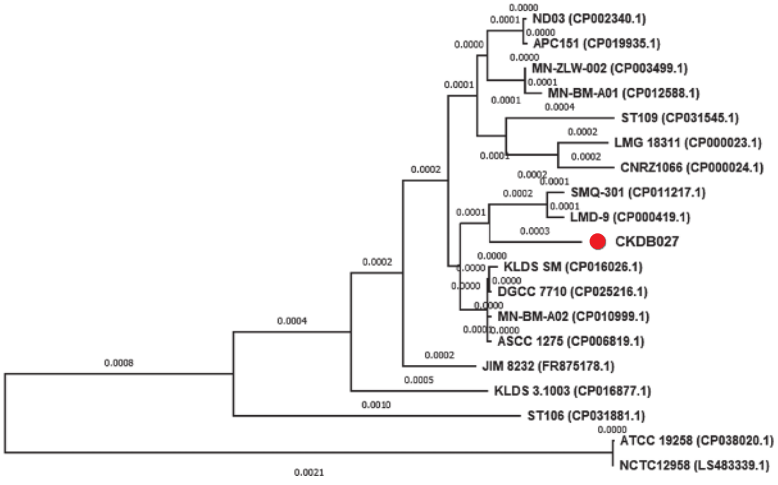


**Fig. 2.** Genome-based phylogenetic tree of CKDB027 with related *S. thermophilus* strains. The phylogenetic tree was built based on the average nucleotide identity values of 19 *S. thermophilus* strains.

[function unknown], 6 category M [cell wall/membrane/envelope biogenesis], 5 category J [translation, ribosomal structure, and biogenesis], 5 category E [amino acid transport and metabolism], 3 category O [post-translational modification, protein turnover, and chaperones], 2 category T [signal transduction mechanisms], 1 category G [carbohydrate transport and metabolism], 1 category K [transcription], 1 category P [inorganic ion transport and metabolism], and 1 category V [defense mechanisms]) (**Fig. 3A**).

The pan- and core-genome analysis involving 19 *S. thermophilus* strains, including CKDB027, showed that the pan-genome contained 1,513 core genomes with 2,866 genes in total (**Fig. 3B**). In addition, from the 16 strains, a total of 318 unique genes (singletons) were found and no unique gene was found in other 3 strains. Based on the result, the b-value of the power-law regression model was calculated as 0.15, indicating that the pan-genome of *S. thermophilus* species is an open genome.[24]


https://doi.org/10.52361/fsbh.2021.1.e28

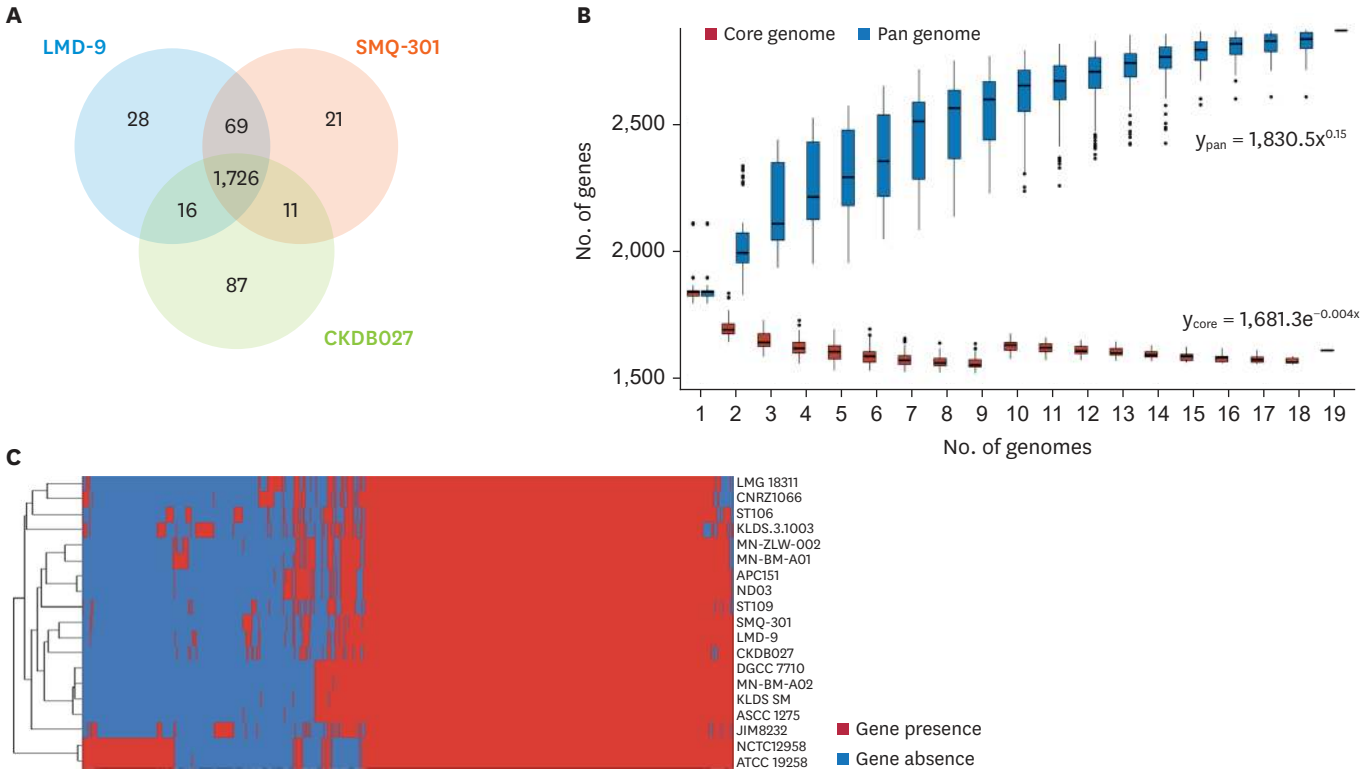


**Fig. 3.** Comparison of functional genes of *S. thermophilus* strains. (A) Venn Diagram analysis of 3 *S. thermophilus* strains LMD-9, SMQ-301 and CKDB027 strain. (B) Pan- and core-genome analysis of *S. thermophilus* strains. (C) Heat-map illustrating the presence and absence of the genes in each genome of the 19 *S. thermophilus* strains.

## Carbohydrate utilization, proteolytic system, and amino acid synthesis

The carbohydrate utilization and proteolytic systems were examined based on gene annotation and comparative genomic analysis. An intact lactose-galactose operon comprising *galR*, *galK*, *galT*, *galE*, *galM*, *lacS*, and *lacZ* was found in the CKDB027 strain, and compared with other *S. thermophilus* strains, a high level of conservation was observed in this strain (**Fig. 4**). The result predicted a lactose degradation ability of the CKDB027 strain.

Strain CKDB027 was shown to have all 12 cytoplasmic peptidase genes: *pepA*, *pepC*, *pepF*, *pepM*, *pepN*, *pepO*, *pepP*, *pepQ*, *pepS*, *pepT*, *pepV*, and *pepX*. In addition, the following genes of the amino acid transporter family that belong to the ATP-binding cassette (ABC) superfamily were found: oligopeptide ABC transporter, branched-chain amino acid ABC transporter, glutamine ABC transporter, spermidine/putrescine ABC transporter, and methionine ABC transporter (**Supplementary Fig. 3**). Interestingly, 9 of the 18 strains of the *S. thermophilus* strains lack *prtS* but exists in the CKDB027 (**Fig. 4**), sharing over 95% conserved with *Streptococcus suis* (data not shown).

In CKDB027, genes encoding 20 amino acid synthesis were also detected (**Fig. 4**, **Supplementary Fig. 4**), but the lack of *dapF* for lysine biosynthesis suggested that lysine biosynthesis is likely to be incomplete in CKDB027. The strain is thus predicted that providing deficient amino acids is depended on other bacteria or the culture media.


https://doi.org/10.52361/fsbh.2021.1.e28

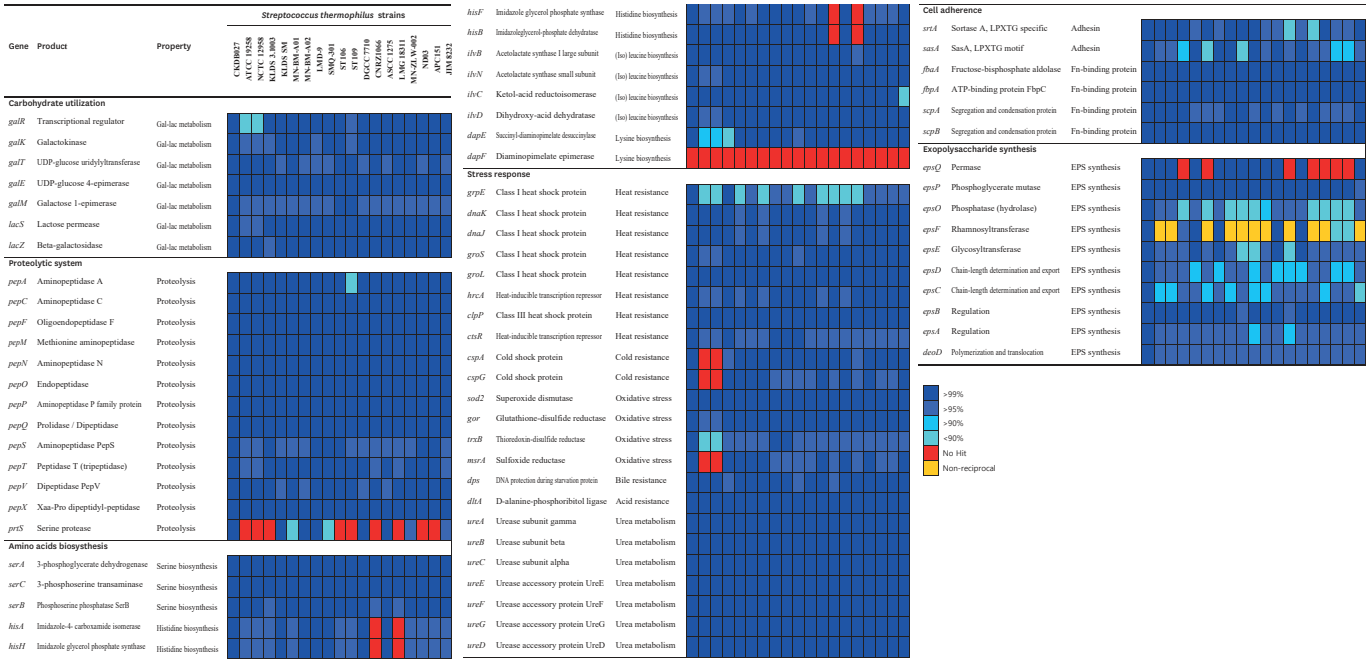


**Fig. 4.** Probiotic property genes in CKDB027 and other *S. thermophilus* strains. Heat map of genes with probiotic functionality among 19 *S. thermophilus* strains used in fermented products and CKDB027. Each color shows the similarity between CKDB027 and other strains.

## Stress resistance

The ability of the bacteria to adapt to low pH, change in osmotic pressure or temperature, and oxidative stress is essential for survival in the human intestine or under industrial process. For the CKDB027 strain, *GroELS* and *grpE-dnaK-dnaJ* encoding the class I heat shock proteins and a negative regulator *hrcA* were detected (**Fig. 4**). The other heat-shock genes included the class III heat shock gene, *clpP*, and its transcription repressor, *ctsR*, along with the cold-shock genes, *cspA*, and *cspG*. The genes related to oxidative stress such as superoxide dismutase, glutathione-disulfide reductase, thioredoxin-disulfide reductase, and sulfoxide reductase were found in CKDB027. In the CKDB027 genome, the genes related to bile resistance and acid resistance were *dps* and *dltA*, respectively, and the genes related to urea metabolism, *ureA*, *ureB*, *ureC*, *ureE*, *ureF*, *ureG*, and *ureD* were also detected (**Fig. 4**). The presence of various genes in association with stress resistance predicted a high level of survival in the human body or under industrial production environment.

## Defense system

The most well-known prokaryotic defense mechanisms include the CRISPR-Cas system, restriction-modification (R-M) system, and bacteriocin production. The whole genome of CKDB027 showed 2 CRISPR-Cas locus clusters (**Fig. 5**): one was the CRISPR-Cas type II-C system with 14 spacers and the other was the CRISPR-Cas type II-A system with 23 spacers. In addition, 1 and 3 spacers were found in different regions, and the gene cluster encoding the components of the type I R-M system was detected. Lastly, the gene cluster producing a type of bacteriocin, the ribosomally synthesized and post-translationally modified peptides (RiPP; streptide, bovicin, and sactipeptides), was also detected (**Supplementary Fig. 5**). The results imply that the CKDB027 strain has an antimicrobial activity with defense system against competing microorganisms in addition to the mechanisms to degrade foreign DNA.


https://doi.org/10.52361/fsbh.2021.1.e28

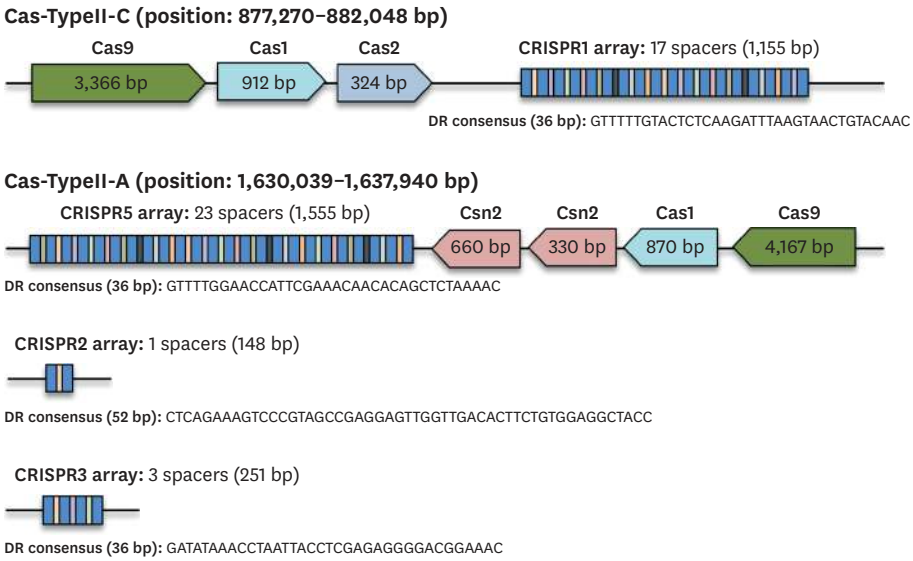


**Fig. 5.** Structure of CRISPR/Cas loci in *S. thermophilus* CKDB027. Four different DRs are blue color-coded and spacers are other color-coded. The consensus sequence and size of each DRs are shown at bottom of each locus. The size of each Cas protein in the locus is indicated in each pentagon.
DR = direct repeat, bp = base pair.

### Cell adherence (host adhesion and interaction)

The CKDB027 genome was shown to contain the genes required for attaching to host epithelial cells. *srtA* and *sasA* which encoding sortase, the most well-known adherence protein found in Gram-positive bacteria, were detected in CKDB027 genome. In addition, the genes encoding fibronectin-binding proteins, *fbaA*, *fbpA*, and *scpAB* which anchor to the host cell through binding to fibronectin on the cell surface, were detected, predicting their assistive role in intestinal adhesion (**Fig. 4**).

### EPS biosynthesis gene clusters

The CKDB027 genome contained the EPS biosynthesis gene cluster related to the production of EPS with a physiological role in enhancing food texture and various bioactivity like host immunity (**Fig. 6**). At the end of the gene cluster, *deoD-epsABCD* and *epsOPQ* were detected, sharing over 95% conserved regions with other strains, whereas the other parts were unique in their composition (**Fig. 4**). The EPS biosynthesis gene cluster of CKDB027 strain also contained 5 glycosyltransferase genes, 1 rhamnosyltransferase gene, 2 chain-length determination gene, 1 UDP-galactopyranose mutase gene, 2 polymerization protein genes, and 2 transposases, predicting a high level of EPS production by the strain.

## DISCUSSION

In this study, whole-genome sequencing and comparative genomic analysis of *S. thermophilus* CKDB027 were performed to identify the phylogenetic position and to validate the safety and probiotic functionality. To perform strain-level identification, we obtained full-length 16S rRNA gene by capillary electrophoresis sequencing (CES) to obtain long-read with low base-calling errors (ABI PRISM 3730XL Analyzer). We performed the BLAST using 16S rRNA gene sequence and identified the CKDB027 strain is similar to *S. thermophilus* species. We conducted whole genome sequencing of CKDB027 strain using PacBio long read sequencer and Illumina short


https://doi.org/10.52361/fsbh.2021.1.e28

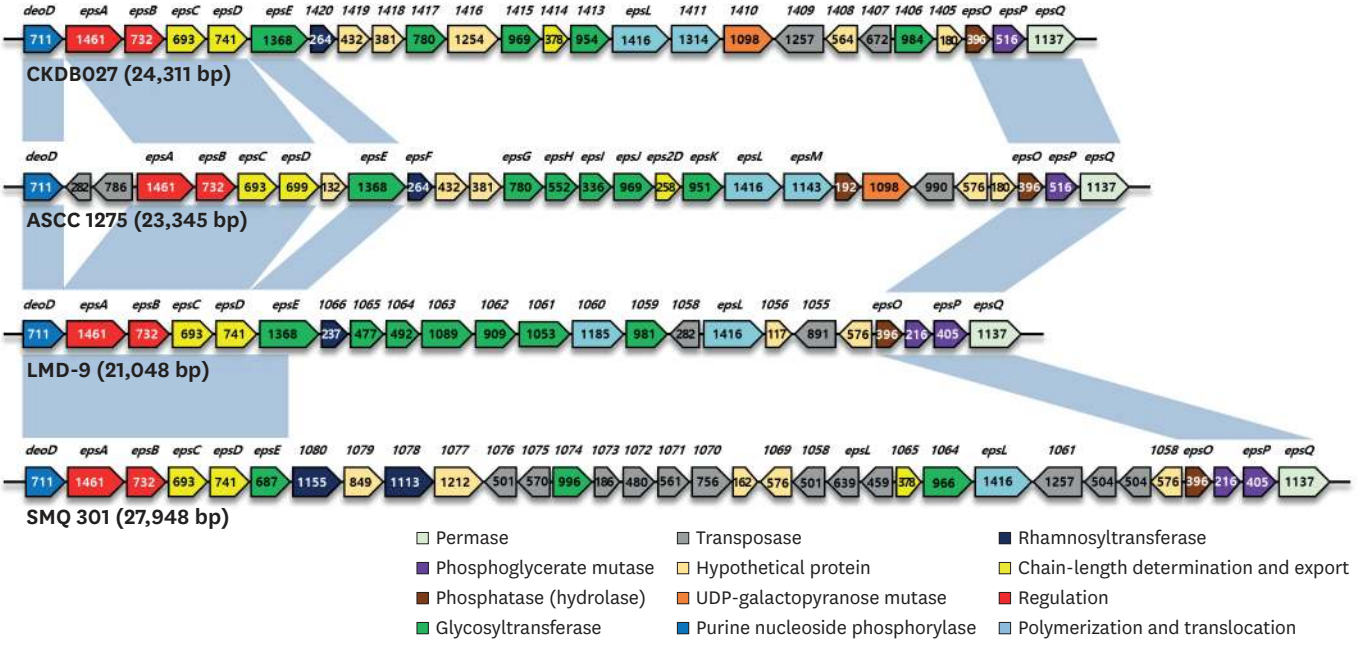


**Fig. 6.** Comparison of EPS gene cluster among *S. thermophilus* CKDB027 and the closely related 4 strains. ORFs were color-coded according to predicted functions indicated at the lower bottom panel. Regions sharing similar sequences among the EPS gene clusters are shown in blue boxes. The size of each gene in the locus is indicated in each pentagon.
EPS = exopolysaccharide, ORF = open reading frame.

read sequencer. We performed completing draft genome through de novo assembly with long reads (10–20kb), followed by correcting minor single nucleotide variation and indels with short reads from Illumina shotgun sequencing. We confirmed sufficient coverage with all regions mapped and minor error through resequencing, so we could assure accuracy of whole genome sequencing. To support accuracy, we also confirmed that the sequence obtained by 16S rRNA sequencing from CES method is identical to the 16S rRNA region sequence from assembled whole genome. These data showed that the CKDB027 strain is identified to *S. thermophilus*. The CKDB027 strain showed similar genomic features to other 18 *S. thermophilus* strains used in dairy products, in terms of genome size and GC content. We identified the genome of CKDB027 strain was similar over 99% with them. The phylogenetic tree through of ANI value showed that the CKDB027 strain was positioned closely to other strains used in fermented products of the *S. thermophilus*. In addition, no genes associated with AR or virulence were detected. The results thus confirmed the CKDB027 strain as a probiotic strain with verified safety. The genetic features of CKDB027 concerning lactose metabolism specialized system, proteolysis system, various stress response and defense mechanisms were similar to those of other related strains. Furthermore, the presence of the EPS gene cluster indicated the possibility of EPS production.

The ANI analysis for the whole genome of the CKDB027 strain and 18 *S. thermophilus* strains used in fermented foods showed that they all shared a common ancestor. Moreover, 87 unique genes that were not shared with other strains were found in the CKDB027 strain; among them, 17 genes were of the COG major category: replication, recombination, and repair.

The pan- and core-genome analysis for 19 *S. thermophilus* strains, including CKDB027 showed that the CKDB027 strain shared 71.2% core genes (n = 1,513) with 18 *S. thermophilus* strains (**Fig. 3B**). By calculating the b-value in the power-law regression model, the pan-genome of the *S. thermophilus* species was identified as an open genome.[24]


https://doi.org/10.52361/fsbh.2021.1.e28

The key probiotic property of *S. thermophilus* is lactose fermentation. Through the Leloir pathway, *S. thermophilus* preferentially ferments lactose over glucose, with an outstanding ability to catabolize lactose.[24] The CKDB027 strain contained *galR*, *galK*, *galT*, *galE*, *galM*, *lacS*, and *lacZ* with a highly conserved lactose-galactose operon as in other *S. thermophilus* strains, which predicts to include lactose catabolism pathway.[41] Another key probiotic property is the presence of a proteolytic system. Casein is hydrolyzed in milk by *S. thermophilus* to obtain amino acids and peptides necessary for growth.[42] One of the core genes in the proteolytic system is *prtS* related in the initiation of the proteolytic cascade and is thus required for the rapid growth of the bacteria in milk.[24] Almost half of the *S. thermophilus* strains lack *prtS*; therefore, a protocooperation with *Lactobacillus delbrueckii* subsp. *bulgaricus* is necessary to induce the proteolytic system.[24] However, the *prtS* gene showing 95% conservation with *S. suis* was found in the CKDB027 strain, suggesting possible growth in milk. The deficient amino acids that could not be obtained through proteolysis should be acquired through amino acid biosynthesis, and in case of CKDB027, among the 20 genes for amino acid synthesis, the *dapF* gene for lysine biosynthesis was missing. This phenomenon is commonly observed across all *S. thermophilus* strains, and the biosynthesis of the missing amino acid is expected to depend on symbiosis with other lactic acid bacteria present in milk.[24]

For the use of lactic acid bacteria in the production of fermented products, such as yogurt and cheese, the ability to survive in the starter culture with unfavorable growth conditions is essential. Moreover, for providing beneficial effect to humans as probiotics, the ability to survive in the gut environment is crucial.[43] Such harsh conditions include low pH, change in osmotic pressure or temperature, and oxidative stress. The CKDB027 strain was shown to have the class I heat shock protein genes (*grpE-dnaK-dnaJ* and *GroELS*) and their negative regulator, *hrc*. In addition, the class III heat shock gene (*clpP*) and its transcription repressor, *ctsR*, were found. The presence of the cold-shock genes (*cspA* and *cspG*) also predicts the heat and cold resistance properties of CKDB027. Based on the detection of genes related to oxidative stress (superoxide dismutase, glutathione-disulfide reductase, thioredoxin-disulfide reductase, and sulfoxide reductase), CKDB027 is also predicted to be able to survive in an acidic environment during fermentation. The presence of the bile resistance gene *dps* and acid resistance gene *dltA* led to the conclusion that survival may be increased in the gut environment.[44] CKDB027 has urea metabolism genes (*ureA*, *ureB*, *ureC*, *ureE*, *ureF*, *ureG*, and *ureD*) to catabolite urea and produce ammonia to increase the pH which affects adversely the acidification rate during fermentation. Such properties are known to be appeared only in the *S. thermophilus* among the genus *Streptococcus*, as an important stress response mechanism harnessed in an acidic environment.[24]

Probiotics are known to be susceptible to phage infection in the process of fermentation or in the intestinal environment,[42,45] therefore, an effective defense mechanism is required. The CRISPR-Cas system is one of defense system that is widely distributed across prokaryotes to provide resistance against foreign factors of a virus or plasmid.[46] We found 2 types of CRISPR-Cas systems (CRISPR-Cas type II-C and CRISPR-Cas type II-A, each containing 14 and 23 spacers, respectively) with complete cluster form, and expected that CKDB027 has potential resistance to phage infection.

Another defense system is the R-M system.[47] The R-M system can also remove foreign DNA, and work complementary to the CRISPR-Cas system. A gene encoding a component of the type I R-M system was found in CKDB027, indicating potential effective defense against foreign DNA. In addition, the gene clusters producing RiPPs, such as streptide, bovicin, and


https://doi.org/10.52361/fsbh.2021.1.e28

sactipeptides, were found. RiPPs play a major role in the antimicrobial activity as well as in the formation of bacterial colonies. Hence, the defense mechanism against foreign agents predicts the competitive superiority of CKDB027 in diverse bacterial communities.

Successful colonization in gut environment is an important property of probiotics.[48] In CKDB027, the genes encoding sortase, *srtA* and *sasA*, were found. These genes produce the representative adherence proteins in Gram-positive bacteria that contain the LPXTG motif (leucine, proline, X, threonine, and glycine, where X is any amino acid) adhered to the cell surface,[42] which allows the bacteria to stably adhere to the intestinal epithelial cells or the mucosa. CKDB027 genome was also found to contain the genes encoding fibronectin-binding proteins, *fbaA*, *fbpA*, and *scpAB*, that bind to the cell surface fibronectin to allow anchoring to the host cell, which predicts continuous interaction with epithelial cells.[49]

EPS are carbohydrate polymers present on the cell surface, whose main function is protection against various environmental stresses. But in fermented foods, such as yogurt, EPS may enhance food texture thus playing a critical role in the dairy industry.[50] EPS is also highlighted as a probiotic functionality by improving human health through various functions, such as by reducing cholesterol levels, enhancing immune system, and exhibiting antiviral and antioxidation effects.[6] In *S. thermophilus*, EPS primarily consist of heterosaccharides, such as galactose, glucose, and rhamnose, which form a gene cluster containing various genes related to glycosyltransferase, acetyltransferase, and polymerization for EPS synthesis.[6] *deoD-epsABCD* and *epsOPQ* of CKDB027 showed over 95% conserved sequence with those of other strains, with variations in other components. In CKDB027 genome, 5 glycosyltransferase genes, 1 rhamnosyltransferase gene, 2 chain-length determination gene, 1 UDP-galactopyranose mutase gene, 2 polymerization protein gene, and 2 transposases were additionally found. All genes related to the production of EPS, such as transferase and polymerization proteins, were found which indicates that CKDB027 have the potential to provide various bioactive functions through the production of EPS.

The results collectively suggest that CKDB027 is a safe lactic acid bacterial strain and does not contain any genes related to AR or virulence. The possibility of various metabolic functions as well as enhanced bioactivities was also demonstrated. These findings indicate that the *S. thermophilus* CKDB027 strain is likely to prove useful in various probiotic industries, including for the manufacturing of fermented products.

## SUPPLEMENTARY MATERIALS

### Supplementary Table 1

Identification of prophage regions in *S. thermophilus* CKDB027 genome

**Click here to view**

### Supplementary Fig. 1

Phylogenetic tree constructed based on the 16S rRNA gene sequences.

**Click here to view**


https://doi.org/10.52361/fsbh.2021.1.e28

**Supplementary Fig. 2**

Genomic islands in *S. thermophilus* CKDB027 genome. From inner to outside of the circle 11 GIs found by SIGI-HMM algorithm (yellow), 14 GIs found by IslandPath-DIMOB algorithm (blue), and all integrated 25 GIs (red).

Click here to view

**Supplementary Fig. 3**

Genes encoding amino acid transporter involved in ABC superfamily in 19 *S. thermophilus* strains.

Click here to view

**Supplementary Fig. 4**

Genes encoding amino acid biosynthesis in 19 *S. thermophilus* strains.

Click here to view

**Supplementary Fig. 5**

Predicted bacteriocin related genes.

Click here to view

## REFERENCES


1. Sanders ME, Merenstein D, Merrifield CA, Hutkins R. Probiotics for human use. Nutr Bull 2018;43(3):212-25. CROSSREF
2. Behnsen J, Deriu E, Sassone-Corsi M, Raffatellu M. Probiotics: properties, examples, and specific applications. Cold Spring Harb Perspect Med 2013;3(3):a010074. PUBMED | CROSSREF
3. Akour A. Probiotics and COVID-19: Is there any link? Lett Appl Microbiol 2020;71(3):229-34. PUBMED | CROSSREF
4. Natarajan P, Parani M. First complete genome sequence of a probiotic *Enterococcus faecium* strain T-110 and its comparative genome analysis with pathogenic and non-pathogenic *Enterococcus faecium* genomes. J Genet Genomics 2015;42(1):43-6. PUBMED | CROSSREF
5. Emamalipour M, Seidi K, Zununi Vahed S, Jahanban-Esfahlan A, Jaymand M, Majdi H, et al. Horizontal gene transfer: from evolutionary flexibility to disease progression. Front Cell Dev Biol 2020;8:229. PUBMED | CROSSREF
6. Wu Q, Tun HM, Leung FC, Shah NP. Genomic insights into high exopolysaccharide-producing dairy starter bacterium *Streptococcus thermophilus* ASCC 1275. Sci Rep 2014;4(1):4974. PUBMED | CROSSREF
7. Cui Y, Xu T, Qu X, Hu T, Jiang X, Zhao C. New insights into various production characteristics of *Streptococcus thermophilus* strains. Int J Mol Sci 2016;17(10):1701. PUBMED | CROSSREF
8. Mitchell TJ. The pathogenesis of streptococcal infections: from tooth decay to meningitis. Nat Rev Microbiol 2003;1(3):219-30. PUBMED | CROSSREF
9. Iyer R, Tomar SK, Maheswari TU, Singh R. *Streptococcus thermophilus* strains: multifunctional lactic acid bacteria. Int Dairy J 2010;20(3):133-41. CROSSREF
10. Mills S, O'Sullivan O, Hill C, Fitzgerald G, Ross RP. The changing face of dairy starter culture research: from genomics to economics. Int J Dairy Technol 2010;63(2):149-70. CROSSREF

11. Flórez AB, Mayo B. Antibiotic resistance-susceptibility profiles of *Streptococcus thermophilus* isolated from raw milk and genome analysis of the genetic basis of acquired resistances. Front Microbiol 2017;8:2608. PUBMED | CROSSREF
12. Kwon G, Heo B, Kwon MJ, Kim I, Chu J, Kim BY, et al. Effect of silk fibroin biomaterial coating on cell viability and intestinal adhesion of probiotic bacteria. J Microbiol Biotechnol 2021;31(4):592-600. PUBMED | CROSSREF
13. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 2017;67(5):1613-7. PUBMED | CROSSREF
14. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 2017;27(5):722-36. PUBMED | CROSSREF
15. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. Forthcoming 2013. https://arxiv.org/abs/1303.3997.
16. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43(5):491-8. PUBMED | CROSSREF
17. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008;9:75. PUBMED | CROSSREF
18. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 2018;46(W1):W246-51. PUBMED | CROSSREF
19. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 2017;45(W1):W30-5. PUBMED | CROSSREF
20. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 2013;57(7):3348-57. PUBMED | CROSSREF
21. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 2005;33(Database issue):D325-8. PUBMED | CROSSREF
22. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res 2018;46(W1):W278-81. PUBMED | CROSSREF
23. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25(17):3389-402. PUBMED | CROSSREF
24. Alexandraki V, Kazou M, Blom J, Pot B, Papadimitriou K, Tsakalidou E. Comparative genomics of *Streptococcus thermophilus* support important traits concerning the evolution, biology and technological properties of the species. Front Microbiol 2019;10:2916. PUBMED | CROSSREF
25. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 2009;106(45):19126-31. PUBMED | CROSSREF
26. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 2018;35(6):1547-9. PUBMED | CROSSREF
27. Linares DM, O'Callaghan TF, O'Connor PM, Ross RP, Stanton C. *Streptococcus thermophilus* APC151 strain is suitable for the manufacture of naturally GABA-enriched bioactive yogurt. Front Microbiol 2016;7:1876. PUBMED | CROSSREF
28. Cho H, Park KE, Kim KS. Genome analysis of *Streptococcus salivarius* subsp. *thermophilus* type strain ATCC 19258 and its comparison to equivalent strain NCTC 12958. Arch Microbiol 2021;203(4):1843-9. PUBMED | CROSSREF
29. Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, Kulakauskas S, et al. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat Biotechnol 2004;22(12):1554-8. PUBMED | CROSSREF

30. Hatmaker EA, Riley LA, O'Dell KB, Papanek B, Graveley BR, Garrett SC, et al. Complete genome sequence of industrial dairy strain *Streptococcus thermophilus* DGCC 7710. Genome Announc 2018;6(6):e01587-17.
PUBMED | CROSSREF
31. Delorme C, Bartholini C, Luraschi M, Pons N, Loux V, Almeida M, et al. Complete genome sequence of the pigmented *Streptococcus thermophilus* strain JIM8232. J Bacteriol 2011;193(19):5581-2.
PUBMED | CROSSREF
32. Evivie SE, Li B, Ding X, Meng Y, Yu S, Du J, et al. Complete genome sequence of *Streptococcus thermophilus* KLDS 3.1003, a strain with high antimicrobial potential against foodborne and vaginal pathogens. Front Microbiol 2017;8:1238.
PUBMED | CROSSREF
33. Li B, Ding X, Evivie SE, Jin D, Meng Y, Huo G, et al. Short communication: genomic and phenotypic analyses of exopolysaccharides produced by *Streptococcus thermophilus* KLDS SM. J Dairy Sci 2018;101(1):106-12.
PUBMED | CROSSREF
34. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, et al. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A 2006;103(42):15611-6.
PUBMED | CROSSREF
35. Bai Y, Sun E, Shi Y, Jiang Y, Chen Y, Liu S, et al. Complete genome sequence of *Streptococcus thermophilus* MN-BM-A01, a strain with high exopolysaccharides production. J Biotechnol 2016;224:45-6.
PUBMED | CROSSREF
36. Shi Y, Chen Y, Li Z, Yang L, Chen W, Mu Z. Complete genome sequence of *Streptococcus thermophilus* MN-BM-A02, a rare strain with a high acid-producing rate and low post-acidification ability. Genome Announc 2015;3(5):e00979-15.
PUBMED | CROSSREF
37. Kang X, Ling N, Sun G, Zhou Q, Zhang L, Sheng Q. Complete genome sequence of *Streptococcus thermophilus* strain MN-ZLW-002. J Bacteriol 2012;194(16):4428-9.
PUBMED | CROSSREF
38. Sun Z, Chen X, Wang J, Zhao W, Shao Y, Wu L, et al. Complete genome sequence of *Streptococcus thermophilus* strain ND03. J Bacteriol 2011;193(3):793-4.
PUBMED | CROSSREF
39. Labrie SJ, Tremblay DM, Plante PL, Wasserscheid J, Dewar K, Corbeil J, et al. Complete genome sequence of *Streptococcus thermophilus* SMQ-301, a model strain for phage-host interactions. Genome Announc 2015;3(3):e00480-15.
PUBMED | CROSSREF
40. Renye JA Jr, Needleman DS, Steinberg DH. Complete genome sequences of bacteriocin-producing *Streptococcus thermophilus* strains ST106 and ST109. Microbiol Resour Announc 2019;8(7):e01336-18.
PUBMED | CROSSREF
41. Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, Leblond-Bourget N, et al. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. FEMS Microbiol Rev 2005;29(3):435-63.
PUBMED
42. Goh YJ, 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 2011;10 Suppl 1(Suppl 1):S22.
PUBMED | CROSSREF
43. Ghattargi VC, Gaikwad MA, Meti BS, Nimonkar YS, Dixit K, Prakash O, et al. Comparative genome analysis reveals key genetic factors associated with probiotic property in *Enterococcus faecium* strains. BMC Genomics 2018;19(1):652.
PUBMED | CROSSREF
44. Ghattargi VC, Nimonkar YS, Burse SA, Davray D, Kumbhare SV, Shetty SA, et al. Genomic and physiological analyses of an indigenous strain, *Enterococcus faecium* 17OM39. Funct Integr Genomics 2018;18(4):385-99.
PUBMED | CROSSREF
45. Ventura M, Sozzi T, Turroni F, Matteuzzi D, van Sinderen D. The impact of bacteriophages on probiotic bacteria and gut microbiota diversity. Genes Nutr 2011;6(3):205-7.
PUBMED | CROSSREF
46. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science 2010;327(5962):167-70.
PUBMED | CROSSREF

47. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 2015;43(Database issue):D298-9.
PUBMED | CROSSREF

48. Alp D, Kuleaşan H. Adhesion mechanisms of lactic acid bacteria: conventional and novel approaches for testing. World J Microbiol Biotechnol 2019;35(10):156.
PUBMED | CROSSREF

49. Han S, Lu Y, Xie J, Fei Y, Zheng G, Wang Z, et al. Probiotic gastrointestinal transit and colonization after oral administration: a long journey. Front Cell Infect Microbiol 2021;11:609722.
PUBMED | CROSSREF

50. Faber EJ, Zoon P, Kamerling JP, Vliegenthart JF. The exopolysaccharides produced by *Streptococcus thermophilus* Rs and Sts have the same repeating unit but differ in viscosity of their milk cultures. Carbohydr Res 1998;310(4):269-76.
PUBMED | CROSSREF