Current Research in Food Science 10 (2025) 101017

Contents lists available at ScienceDirect

Current Research in Food Science

journal homepage: www.sciencedirect.com/journal/current-research-in-food-science

Multi-omics analysis reveals the mechanism for galactose metabolism in mutant *Streptococcus thermophilus* IMAU20551Y

Jiahui Tai [a,b,c], Haimin Hu [a,b,c], Jinhui Liu [a,b,c], Wenhui Lu [a,b,c], Tong Dan [a,b,c,*]

[a] Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education, Inner Mongolia Agricultural University, Hohhot, China
[b] Key Laboratory of Dairy Products Processing, Ministry of Agriculture and Rural Affairs, Inner Mongolia Agricultural University, Hohhot, China
[c] Inner Mongolia Key Laboratory of Dairy Biotechnology and Engineering, Inner Mongolia Agricultural University, Hohhot, 010018, China

ARTICLE INFO

Keywords:
Streptococcus thermophilus
Galactose metabolism
Genomic analysis
Transcriptome analysis

ABSTRACT

Streptococcus thermophilus (*S. thermophilus*) is a species widely used in the dairy industry to accelerate the acidification rate and improve the texture and flavour characteristics of dairy products. However, most *S. thermophilus* have galactose-negative (Gal^-) phenotypes, which can lead to accumulation of free galactose in fermented dairy products. In a previous study, a mutant of *S. thermophilus* IMAU20551Y was obtained by N-methyl-N′-nitro-N-nitrosoguanidine (NTG) mutagenesis in which key enzymes related to galactose metabolism were significantly changed compared with the wild type. β-galactosidase and galactokinase activity were higher in the mutant while glucokinase and pyruvate kinase activities were significantly decreased compared with the wild type. In this study, the ability of the mutant to metabolize galactose was verified by high performance liquid chromatography (HPLC), and the mechanism for enhanced galactose metabolism elucidated by multi-omics analysis. HPLC analysis showed that accumulation of galactose in milk fermented by mutant *S. thermophilus* IMAU20551Y was reduced by 41.4%, compared with the wild type. Although no mutations in gene sequences associated with galactose metabolism were detected by genome sequencing, transcriptomic data showed up-regulation in expression of *galM*, *galK*, *galT*, *galE* (associated with the Leloir pathway) and LacI family transcriptional regulator GalR, resulting in enhanced galactose metabolism in the mutant. This study provides a reference for genetic engineering modification of galactose-positive (Gal^+) *S. thermophilus*, which is expected to be used as a starter for the production of low galactose fermented dairy products.

1. Introduction

Streptococcus thermophilus (*S. thermophilus*) is an important homofermentative lactic acid bacterium (LAB) that metabolizes carbohydrates such as lactose and glucose into lactic acid, and is widely used in the production of fermented dairy products (Xu et al., 2022). Lactose is the main carbon source in milk and is formed by linking two monosaccharides, α/β-d-glucose and β-d-galactose, through β-1,4 glycosidic bonds (Iskandar et al., 2019). When lactose enters LAB, it is catalyzed by β-Gal to rapidly produce glucose and galactose, the glucose enters the glycolysis pathway, and the galactose is metabolized by the Leloir pathway (Conte et al., 2021). However, most *S. thermophilus* exhibit a galactose-negative (Gal^-) phenotype and have lower galactokinase activity than galactose-positive (Gal^+) strains, which was attributed to poorer translation of the *galK* gene in Gal^- *S. thermophilus* (Cui et al., 2016). This metabolic defect results in accumulation of free galactose in fermented dairy products, which has adverse effects on the product (e.g. defects in cheese texture, browning, microbial spoilage), and on human health (e.g. toxicity effects in patients with galactosemia) (Wu et al., 2015). Johnson and Olson (1985) reported that large amounts of galactose remained in mozzarella cheese when Gal^- *S. thermophilus* strains were used in production, and that the amount of galactose was positively correlated with the degree of cheese browning when it was heated. Therefore, the ability to use galactose is a desirable characteristic in strains intended for industrial production of fermented dairy products.

Given the defective galactose metabolism of *S. thermophilus*, genetic improvement of *S. thermophilus* is a priority. Mutation breeding is one of

* Corresponding author. Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education, Inner Mongolia Agricultural University, Hohhot, China.
E-mail addresses: Tai_1888@163.com (J. Tai), huhaimin99@163.com (H. Hu), 15225176844@163.com (J. Liu), 15248180034@163.com (W. Lu), dantong813218@aliyun.com (T. Dan).

https://doi.org/10.1016/j.crfs.2025.101017
Received 19 September 2024; Received in revised form 26 February 2025; Accepted 1 March 2025
Available online 2 March 2025

the most promising approaches for increasing genetic and phenotypic variability (Shahwar et al., 2023). Chemical mutagenesis has been widely used in bacteria because of its low cost, convenient use and strong specificity (Hu et al., 2020). NTG is a chemical super-mutagen which interacts with nucleic acid bases causing changes in the structure of DNA molecules, while also hindering the unravelling of DNA double strands, thus leading to gene mutations (Harper and Lee, 2012). Benateya et al. (1991) used NTG to treat *S. thermophilus* CNRZ302, and its A5 mutant strain was able to metabolize 70% of the galactose produced by lactose hydrolysis while also retaining other physiological characteristics. Hu et al. (2024) used NTG to mutate *S. thermophilus* and screened for a Gal $^{+}$ mutant strain that significantly altered metabolites secreted into the culture medium.

Current studies largely focus on mutagenesis screening of Gal $^{+}$ mutants of *S. thermophilus*, but the molecular mechanisms targeting gene expression in the relevant metabolic pathways of Gal^{+} mutants are not yet clear. Hence, in this study, we used multi-omics techniques to investigate the mechanism for enhanced galactose metabolism in Gal $^{+}$ mutant *S. thermophilus* IMAU20551Y. Genetic polymorphisms represent a significant factor influencing phenotypic variation, and their interaction with the environment is of paramount importance in determining the expression of individual traits (Potts and Hunter, 2021). Next-generation sequencing technology is one of the main means of genomics research; it can compare individual genome sequences with reference genome sequences to screen out single nucleotide polymorphic sites (SNPs) and insertion deletion mutations (InDels) (Satam et al., 2023). Transcriptomics is a new branch of functional genomics that provides a precise and comprehensive account of changes in the level of transcription of genes in the cell, and is able to annotate all types of RNAs, including non-coding small RNAs (Mathlin et al., 2020). Sun et al. (2018) combined genomic and transcriptomic analyses to reveal that split and differentially-expressed genes involved in the pyruvate metabolic pathway may be responsible for increased L-lactic acid production by *Lactobacillus rhamnosus* SCT-10-10-60. Yang et al. (2024) used genomic and transcriptomic analyses to determine that up/down-regulation of genes associated with glycolysis/gluconeogenesis suggested a shift in glucose distribution favoring the pentose phosphate pathway and positively influencing GSH production.

In a previous study, the mutant *S. thermophilus* IMAU20551Y was produced using N-methyl-N′-nitro-N-nitrosoguanidine (NTG) to treat *S. thermophilus* IMAU20551. β-galactosidase and galactokinase activity were higher in the mutant while glucokinase and pyruvate kinase activities were significantly decreased compared with the wild type (Data not published). However, whether the mutant is able to metabolize more galactose; and the molecular mechanisms of targeted gene expression in related metabolic pathways, remain unclear. Therefore, in this study, differences in metabolic levels between the wild type *S. thermophilus* IMAU20551 and the mutant *S. thermophilus* IMAU20551Y were determined by high performance liquid chromatography (HPLC), and the mechanism for enhanced galactose metabolism in the mutant was elucidated by combining genomics and transcriptomics. This study lays a new theoretical and practical foundation for further exploration of the regulatory mechanisms in Gal $^{+}$ *S. thermophilus* and provides a reference for the use of genetic engineering to modify Gal $^{+}$ *S. thermophilus*.

2. Material and methods

2.1. Strains, reagents and culture conditions

S. thermophilus IMAU20551 was isolated from traditional fermented yoghurt produced in Zabkhan Province, Mongolia, and provided by the National Collection of Microbial Resource for Feed, Key Laboratory of the Ministry of Education of Dairy Biotechnology and Engineering, Inner Mongolia Agricultural University. Mutant *S. thermophilus* IMAU20551Y was obtained from wild type *S. thermophilus* IMAU20551 treated by NTG mutagenesis. *S. thermophilus* IMAU20551 and IMAU20551Y were cultured at 42 °C for 18 h in M17 broth (Baiwei Biotech Co. Ltd., Guangdong, China). Methanol and acetonitrile were chromatographically pure and purchased from Thermo Fisher Scientific (Shanghai, China). Lactose, galactose, glucose and lactic acid standards (chromatographic grade) were purchased from Yuanye Bio-Technology Co., Ltd (Shanghai, China).

2.2. Metabolite analysis by HPLC

2.2.1. Fermented milk preparation

Distilled water was preheated to 60 °C, and whole milk powder (11.5%) was added and stirred to combine. Once the milk powder was mixed evenly, sucrose (0.1%) was added and heated at 65 °C for 30 min. Homogenisation was performed twice consecutively under conditions of low-pressure 15 MPa and high-pressure 35 MPa. The milk was sterilised at 95 °C for 5 min and then rapidly cooled and stored at 4 °C for later use. The *S. thermophilus* IMAU20551 and its mutant *S. thermophilus* IMAU20551Y were inoculated into whole milk at an inoculum rate of 5 $\times$ 10^{7} (CFU)/mL and fermented at 42 °C. Samples were taken at 0, 2, 4, and 6 h of fermentation and at 0, 1, 3, 7, and 14 d of storage for testing, and three samples were taken at each time point.

2.2.2. Determination of lactose, galactose and glucose in fermented milk by HPLC

Samples of shaken fermented milk (1.0 g) were diluted in ultra-pure water to achieve 25 mL, and then shaken again to mix. Each sample was ultrasonicated for 30 min, centrifuged at 8000×*g* for 15 min, and the supernatant filtered through a 0.45 μm luer syringe filter. A refractive index detector (1260 RID, Agilent Technologies, CA, USA) was used to quantify sugar content, with an injection size of 10 μL for each analysis. The chromatographic separation was done on an Agilent amino column (5 μm, 250 mm × 4.6 mm) at 35 °C. The mobile phase was a mixture of acetonitrile and ultra-pure water with a volume ratio of 70∶30. The elution flow rate was 1.0 mL/min.

2.2.3. Determination of lactic acid in fermented milk by HPLC

Samples of shaken fermented milk (0.5 g) were diluted in 0.01 mol/L H_2SO_4 to achieve 10 mL, mixed well, sonicated for 30 min, centrifuged (8000×*g*, 15 min) and the supernatant filtered through 0.45 μm luer syringe filter to obtain the sample solution to be tested. Lactic acid was detected using an ultraviolet detector with a detection wavelength of 210 (Agilent Technologies, USA). The chromatographic column was a Zorbax SB-Aq column (5 μm, 4.6 mm × 150 mm) at 35 °C. The mixture for mobile phase was 0.01 mol/L phosphate buffer (pH = 2.0) and methanol, the volume ratio was 97∶3, the sample size was 10 μL, and the elution flow rate was 0.5 mL/min.

2.3. Genome sequencing

2.3.1. Genome sequencing, assembly and functional annotation

Bacterial genome *de novo* sequencing was done on the wild type IMAU20551 and mutant strain IMAU20551Y. Both strains were incubated in M17 broth at 42 °C until mid-logarithmic growth was achieved, and the cell pellet collected after washing with PBS buffer. The cell pellet was sent to Majorbio Bio-Pharm Technology Co. Ltd. (Shanghai, China) for sequencing using the Illumina Hiseq2000 platform. Genomic DNA from the two strains was extracted using a TruSeq Nano DNA LT Sample Preparation Kit (Illumina, San. Diego, California, USA). The quality of extracted DNA was examined by agarose gel electrophoresis and quantified using a Qubit® 2.0 fluorometer (DY001422, Thermo Scientific Co., USA) to ensure that extracted DNA met the requirements for subsequent testing. Genomic DNA was fragmented by a Covaris setting (Covaris, S220, Woburn, Massachusetts, USA). DNA fragments with a size of 350 bp were selected to construct genome sequencing libraries. After sequencing by bridge PCR, paired-end sequencing was done using the Illumina Hiseq2000 sequencing platform. Clean data that met the

requirements after quality control were assembled from scratch by SOAPdenovo2 assembler to obtain the genome sequence (Luo et al., 2012).

Coding sequences in the genomes were predicted by Prodigal. Nucleic acid sequences and amino acid sequences of functional genes were obtained by prediction. The tRNAs and rRNAs contained in genomes were predicted by tRNAscan-SE v2.0 and Barrnap v0.8 software, respectively (Kalvari et al., 2018). The predicted genes were functionally annotated through COG (Cluster of Orthologous Groups of proteins, http://www.ncbi.nlm.nih.gov/COG/) and KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/) databases. Genome visualization was by CGView (Stothard et al., 2019).

2.3.2. Comparative genomic variation analysis

Genomes of the wild type IMAU20551 and the mutant strain IMAU20551Y were examined using MUMmer (Version 3.23, http://mummer.sourceforge.net/) comparison software to identify regions of large-scale co-linearity between the genomes; the show-snps function was used to generate SNPs and small InDels (<50 bp) (Kosugi and Terao, 2024). SnpEff (http://snpeff.sourceforge.net/SnpEff.html) software was used for annotation of SNP and InDel sites (Cingolani et al., 2012). The SNP and InDel mutated genes were further annotated by COG and KEGG databases.

2.4. Transcriptome analysis

2.4.1. RNA extraction

Wild type IMAU20551 and mutant strain IMAU20551Y were cultured in M17 liquid medium at 42 °C until the logarithmic phase was achieved, and the cell pellet was collected by centrifugation (8000 × *g*, 10 min, 4 °C) and washed twice with PBS buffer, then quickly frozen with liquid nitrogen. Total RNA was extracted from samples using TRIzol Reagent (Invitrogen, CA, USA). Quality of extracted RNA was assessed using an Agilent 2100 Bioanalyzer, and its concentration and purity was tested using a Nanodrop2000. High quality RNA samples were used to construct cDNA libraries. The total amount of RNA required for a single library was 2 μg, the RNA integrity number was ≥7.0, the concentration was ≥100 ng/μL and the $OD_{260/280}$ was between 1.8 and 2.2.

2.4.2. cDNA library preparation and transcriptome sequencing

cDNA libraries were constructed using TruSeqTM Stranded Total RNA Library Prep Kit (Illumina Inc., San Diego, CA, USA). rRNA was removed from samples using the RiboCop rRNA Removal Kit (Lexogen, Inc., Greenland, NH, USA) following the manufacturer's procedure. mRNA was randomly broken into small fragments of about 200bp by adding fragmentation buffer. The sheared mRNA was used as a template to synthesize the first strand cDNA using random primers in reverse. The second strand cDNA was synthesised while replacing dTTP with dUTP and then digested with uracil-N-glycosylase enzyme so that the library contained only the first strand cDNA. RNA-seq libraries were sequenced on an NovaSeqXPlus platform at Majorbio Bio-pharm Technology Co., Ltd (Shanghai, China). Raw data was then filtered to obtain high quality data. Clean data after quality control was compared with the reference genome to obtain mapped data for subsequent analyses.

2.4.3. Analysis of differentially-expressed genes

Differential gene expression in the mutant was identified by comparing the gene dataset of the wild type with that of the mutant. Differential gene expression between groups was analyzed using DESeq2 in the R package. Differentially-expressed genes between the comparison groups were obtained according to the screening conditions of p-adjust <0.05 and $|\log_2 FC| \geq 1$. The software Goatools was used for GO enrichment analysis of the screened differential genes. A significant enrichment of GO function was confirmed when FDR <0.05. Similar to GO enrichment analysis, KEGG Pathway enrichment analysis of differentially-expressed genes in the gene set was performed using R scripts.

2.5. Statistical analysis

Each experiment had three replications (n = 3) and data were expressed as mean ± standard deviation. Excel and Origin 9.1 were used for data processing and analysis. One-way ANOVA was performed using IBM SPSS Statistics 23 statistical software. The statistical significance level was $P < 0.05$.

3. Results

3.1. Analysis of key metabolites in fermented milk of S. thermophilus IMAU20551 and its mutant

In order to further verify the mutant's ability to metabolize galactose, lactose, galactose, glucose and lactic acid concentrations in milk fermented by *S. thermophilus* IMAU20551 and its mutant were determined by HPLC at different times during fermentation (0, 2, 4, 6 h) and storage (0, 1, 3, 7, 14 d) (Table 1). The lactose concentration in milk fermented by *S. thermophilus* IMAU20551 and its mutant strain was 136.72 mM and 137.60 mM at 0 h of fermentation, and 80.63 mM and 63.69 mM after 14 d of storage, respectively. Compared with *S. thermophilus* IMAU20551, the ability of the mutant to metabolize lactose was increased by 12.69%.

The concentration of galactose in milk fermented by *S. thermophilus* IMAU20551 increased gradually with extension of fermentation and storage time reaching 32.19 mM after 14 d storage. However, the galactose content of the milk fermented by the mutant was 18.87 mM after 14 d of storage, and its galactose accumulation was significantly lower than that of *S. thermophilus* IMAU20551 ($P < 0.05$).

The glucose concentration of milk fermented by *S. thermophilus* IMAU20551 was detected only after 2, 4 and 6 h of fermentation (6.11–12.21 mM), but was not detected during storage, which may be because glucose was metabolized through glycolysis during fermentation, and the detection limit was not reached. However, free glucose was detected in milk fermented by mutant strain IMAU20551Y after 14 d of storage, which indicated that the metabolic pathway of the strain might have been changed by mutagenesis, and more glucose could be secreted into the fermented milk.

The lactic acid concentration of milk fermented by *S. thermophilus* IMAU20551 and its mutant strain increased from 7.44 to 6.22 mM to 92.81 and 102.01 mM, respectively, during fermentation and storage. Compared with *S. thermophilus* IMAU20551, the lactic acid concentration of milk fermented milk by the mutant increased by 12.22%, indicating that the mutant had a stronger acid production ability.

3.2. Identification of genomic changes by genome sequencing

3.2.1. Genomic characterisation of S. thermophilus IMAU20551 and IMAU20551Y

The whole genome of *S. thermophilus* IMAU220551 and its mutant strain IMAU220551Y were re-sequenced. After assembling compliant Clean Reads, the genomic information of the two strains was compared (Table 2). Both IMAU20551 and IMAU20551Y had a genome size of 1.74 Mb and a GC content of 39.07, with no significant differences between the two strains. A total of 1876 CDSs were predicted in IMAU20551. Compared with the wild type IMAU20551, the genome of IMAU20551Y contains 5 more CDSs.

3.2.2. Genomic functional annotation of S. thermophilus IMAU20551 and IMAU20551Y

COG database comparison allows functional annotation and categorisation of predicted proteins as well as evolutionary analysis of proteins. The coding genes predicted by IMAU20551 and IMAU20551Y

Table 1
Lactose, galactose, glucose and lactic acid concentration in milk fermented by *S. thermophilus* IMAU20551 and its mutant.

Time	Lactose (mM)		Galactose (mM)		Glucose (mM)		Lactic acid (mM)	
	IMAU20551	IMAU20551Y	IMAU20551	IMAU20551Y	IMAU20551	IMAU20551Y	IMAU20551	IMAU20551Y
0 h	136.72 ± 2.25^{a}	137.60 ± 0.50^{a}	–	–	–	–	–	–
2 h	129.71 ± 3.71^{a}	130.59 ± 1.26^{a}	7.77 ± 0.11^{a}	8.33 ± 0.22^{a}	6.11 ± 0.22^{b}	12.77 ± 0.11^{a}	7.44 ± 0.36^{a}	6.22 ± 0.06^{b}
4 h	118.90 ± 2.28^{a}	114.81 ± 2.69^{a}	12.77 ± 0.78^{b}	15.54 ± 0.44^{a}	14.43 ± 0.17^{b}	21.09 ± 0.56^{a}	20.32 ± 0.16^{b}	25.31 ± 0.08^{a}
6 h	104.59 ± 0.70^{a}	96.70 ± 0.82^{b}	22.20 ± 0.72^{b}	27.20 ± 0.44^{a}	12.21 ± 0.39^{b}	26.09 ± 0.39^{a}	42.30 ± 0.07^{b}	50.51 ± 0.10^{a}
0 d	95.24 ± 1.61^{a}	84.43 ± 1.75^{b}	26.64 ± 0.39^{a}	26.09 ± 0.28^{a}	–	23.87 ± 0.06	62.83 ± 0.04^{b}	71.60 ± 0.19^{a}
1 d	92.32 ± 1.26^{a}	77.42 ± 0.50^{b}	28.86 ± 0.44^{a}	23.31 ± 0.17^{b}	–	22.76 ± 0.33	73.27 ± 0.19^{b}	86.37 ± 0.18^{a}
3 d	85.89 ± 0.41^{a}	69.82 ± 2.02^{b}	30.53 ± 0.94^{a}	21.65 ± 1.17^{b}	–	21.09 ± 0.22	83.81 ± 0.12^{b}	94.92 ± 0.04^{a}
7 d	82.09 ± 1.37^{a}	65.44 ± 3.07^{b}	31.08 ± 0.38^{a}	21.09 ± 0.67^{b}	–	20.54 ± 0.33	91.36 ± 0.26^{b}	100.91 ± 0.13^{a}
14 d	80.63 ± 1.29^{a}	63.69 ± 1.05^{b}	32.19 ± 0.56^{a}	18.87 ± 0.61^{b}		19.98 ± 0.39	92.81 ± 0.11^{b}	102.01 ± 0.12^{a}

Note: indicates no detection; For each sugar, values in the same row (i.e. taken at the same time) followed by different lower case letters indicate the values are significantly different to each other ($P < 0.05$).

Table 2
Genomic characterisation of *S. thermophilus* IMAU20551 and IMAU20551Y.

Sample Name	*S. thermophilus* IMAU20551	*S. thermophilus* IMAU20551Y
Genome Size (bp)	1,740,218	1,739,718
GC Content (%)	39.07	39.07
Depth	699.36	789.67
CDS No.	1876	1881
tRNA No.	37	41
rRNA No.	2	2

were compared and annotated with the COG database (Fig. 1). There were 1589 CDSs annotated to COG function in the genome of IMAU20551 and 1591 CDSs in that of IMAU20551Y. The CDSs of both strains could be classified into 23 functional categories. The major functional categories were related to translation, ribosomal structure and biogenesis, amino acid transport and metabolism, cell wall/membrane/envelope biogenesis, carbohydrate transport and metabolism and coenzyme transport and metabolism. In the genome of IMAU20551Y, there are 2 more CDSs than in the wild type strain IMAU20551, which were distributed in the category of translation, ribosomal structure and biogenesis, cell wall/membrane/envelope biogenesis.

Functional annotation of IMAU20551 and IMAU20551Y coding genes predicted by the KEGG database (Fig. 2). There were 1759 genes annotated to KEGG function in IMAU20551. KEGG annotations classify genes into 6 major categories, including cellular processes (97), environmental information processing (194), genetic information processing (157), human diseases (117), metabolism (1171), and organismal systems (23). There were 1763 genes annotated to KEGG function in IMAU20551Y, including cellular processes (98), environmental information processing (195), genetic information processing (159), and the number of genes annotated to other pathways is the same as IMAU20551. These differentially-functional genes that were annotated may contribute to the difference in physiological function between IMAU20551 and IMAU20551Y.

3.2.3. SNPs and InDels in comparison to the wild type IMAU20551 genome

In order to determine whether NTG mutagenesis caused genetic changes in the mutant strain, we compared the genome sequences of the two strains using the wild type IMAU20551 as the reference genome and the mutant strain IMAU20551Y as the comparative genome. Compared with IMAU20551, there were 21 SNPs and 22 InDels (12 deletions and ten insertions) detected in the genome of IMAU20551Y. There were 14 SNPs and 19 InDels located in the CDS region. There were seven SNPs and three InDels located in the intergenic region. Using the reference sequence as a scale, Circos (Version0.64) software was used to display the distribution of genes in SNP and InDel on the ring map, so that the distribution of variation analysis results could be visually compared (Fig. 3).

Genetic variants occurring in the CDS region may result in altered gene functions. Therefore, we further analyzed variation in SNPs and InDels occurring in the CDS region. Mutant genes involved in 14 SNPs and 19 InDels were mapped to the COG and KEGG databases, respectively. A total of two genes were annotated to COG function, encoding the magnesium and cobalt transport protein CorA (gene0923) and the ABC transporter (gene0617) (Table 3). In total, five genes were annotated to KEGG function, including genes encoding enolase (gene0274), spermidine/putrescine transport system substrate-binding protein (gene0958), D-methionine transport system permease protein (gene0617), peptidoglycan DL-endopeptidase CwlO (gene0086) and putative transposase (gene1653) (Table 3). These mutated genes may have an impact on different aspects of cell function.

3.3. Transcriptome analysis

3.3.1. Quality assessment of extracted RNA and sequencing data

Total RNA was extracted from wild type *S. thermophilus* (IMAU20551_1, IMAU20551_2 and IMAU20551_3) and mutant strain *S. thermophilus* (IMAU20551Y_1, IMAU20551Y_2 and IMAU20551Y_3), respectively. The RNA integrity number, $OD_{260/280}$, and concentration of six samples were ≥7.0, ≥1.8 and ≥100 ng/μL, respectively (Table S1). All samples met the requirements for the construction of RNA-Seq libraries. RNA-Seq libraries were constructed with the qualified RNA samples to obtain their whole transcriptomes. A total of 14,138,068 clean reads were obtained from six transcriptome sequencing libraries. The number of clean reads was large, the Q20 and Q30 were both greater than 90%, and sequencing error rate was less than 0.1% (Table S2). The above results indicate that the sequencing data of all samples was high quality and suitable for subsequent analyses.

3.3.2. GO and KEGG enrichment analysis of differentially-expressed genes

The DESeq2 in the R package was used to identify differentially-expressed genes. Differentially-expressed genes were screened according to $|\log_2 FC| \geq 1$ and p-adjust <0.05, and the results were visualised by volcano plot (Fig. 4). Compared with wild type IMAU20551, the mutant strain IMAU20551Y had a total of 376 differentially-expressed genes, of which 195 genes were up-regulated and 181 genes were down-regulated. After this, we performed a cluster analysis of gene expression patterns using the H-cluster method and visualised the results by heatmap (Fig. 5). Results showed that differentially-expressed genes were similar across the biological repeats of each strain, but varied widely between strains.

To investigate the specific functional changes in the mutant strain after mutagenesis, we performed GO and KEGG functional enrichment analysis of differentially-expressed genes (Fig. 6). The results of GO enrichment analysis showed that there were 89 differentially-expressed genes annotated into the GO terms, all of which were distributed in biological processes and molecular functions. The largest number of differentially-expressed genes were associated with the terms, viz. galactose metabolic process, amine metabolic process, ATPase-coupled

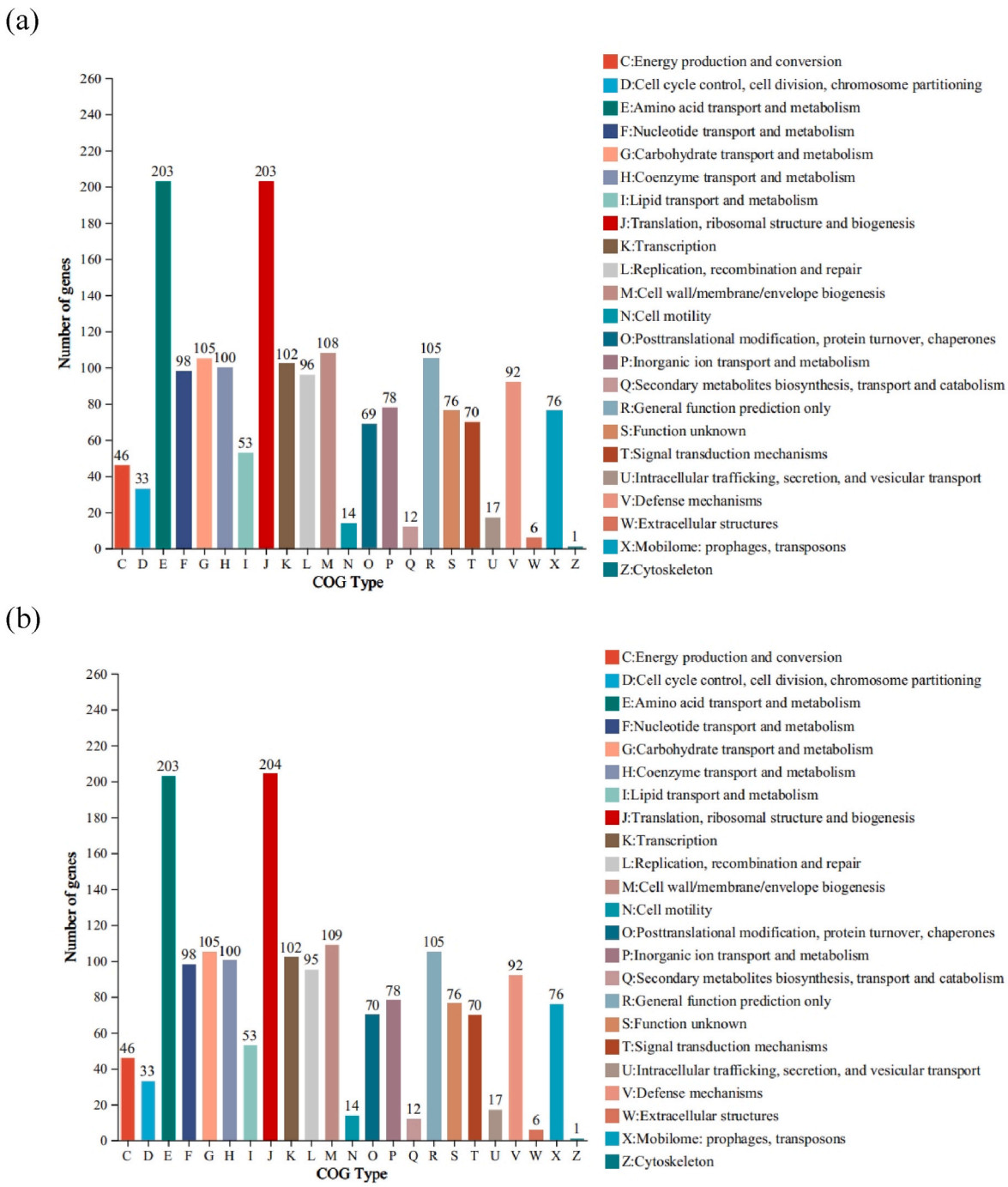


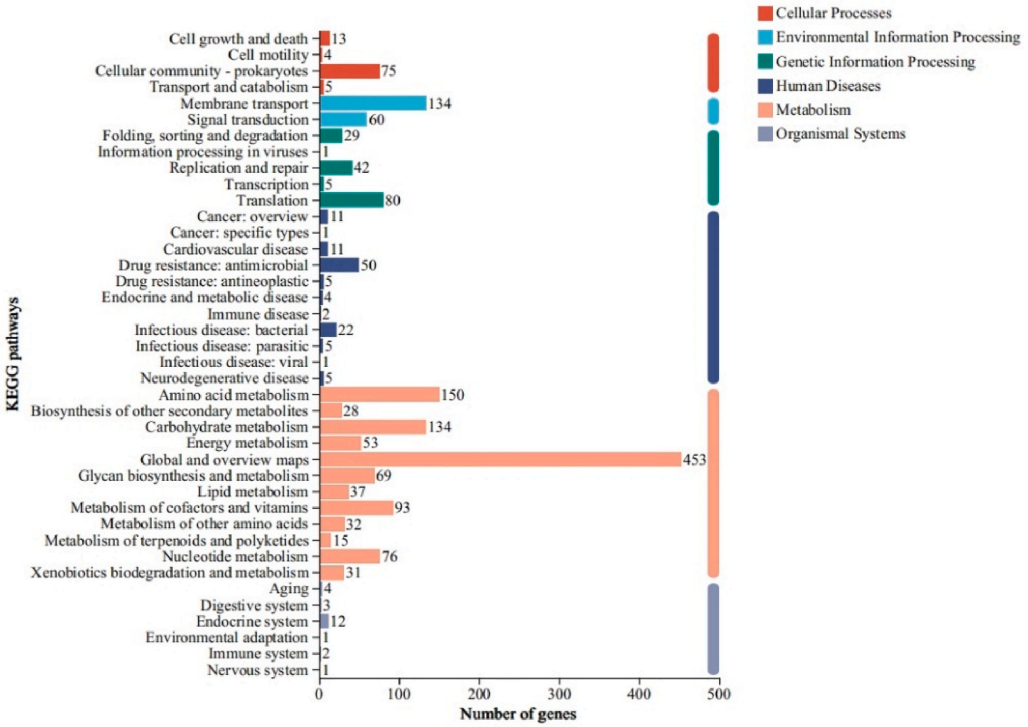
Fig. 1. COG functional classification of *S. thermophilus* IMAU20551 (a). COG functional classification of *S. thermophilus* IMAU20551Y (b). Note: The horizontal axis represents different COG types, and the vertical axis represents the number of genes.

phosphate ion transmembrane transporter activity, phosphate ion transmembrane transporter activity and ATPase-coupled inorganic anion transmembrane transporter activity. KEGG pathway enrichment analysis was also done, and showed 90 differentially-expressed genes were enriched in the KEGG pathway. These differentially-expressed genes were mainly involved in biosynthesis of nucleotide sugars (six), amino sugar and nucleotide sugar metabolism (six), o-antigen nucleotide sugar biosynthesis (four), pyrimidine metabolism (seven), galactose metabolism (five), among others. Combined with the results of GO and KEGG enrichment analysis, we found that the differentially-expressed genes related to carbohydrate metabolism and transporter activity in mutant which could influence phenotypic characteristics. Next, the differentially-expressed genes annotated during galactose and glycolytic metabolism were analyzed to explore the internal mechanism for phenotypic change in mutant IMAU20551Y.

3.3.3. *Differential gene expression related to galactose metabolism*

Galactose is catabolized mainly through the Leloir pathway. Four genes (*galK*, *galT*, *galE*, *galM*) located upstream of the lac operon in *S. thermophilus* are associated with galactose metabolism. Transcriptomic results showed that gene1073 (*galM*), gene1070 (*galK*), gene1071 (*galT*) and gene1683 (*galE*) were all up-regulated in mutant IMAU20551Y compared with wild type (Fig. 5). In addition, expression of gene1069 (*lacI*), which encodes the LacI family DNA-binding

(a)



(b)

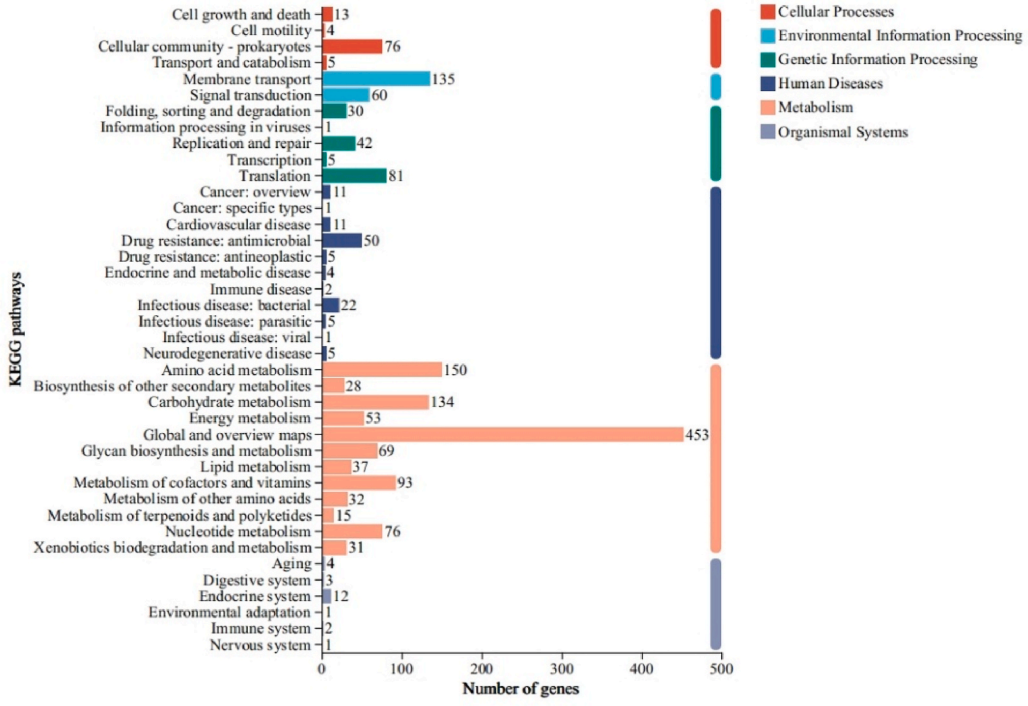


Fig. 2. KEGG functional classification of *S. thermophilus* IMAU20551 (a). KEGG functional classification of *S. thermophilus* IMAU20551Y (b). Note: The vertical axis is the level2 classification of the KEGG pathway, and the horizontal axis is the number of genes annotated under that classification. The column colours represent the level1 classification of the KEGG pathway.

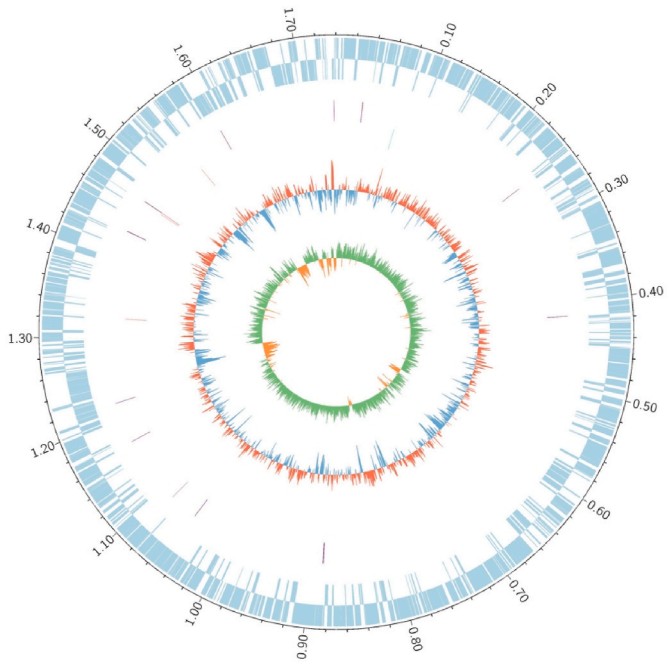


Fig. 3. The genomic variation Circos map. Note: Starting from the outside, the following items are shown: Gene location on the reference genome, CDS on the positive chain, CDS on the negative chain, SNP distribution of the sample, InDel distribution of the sample, GC content of the reference genome.

transcriptional regulator GalR, was up-regulated in the mutant strain compared with the wild type (Fig. 5). Up-regulation of expression of all the above genes promoted metabolism of galactose in the mutant strain.

3.3.4. Differential gene expression related to the glycolysis pathway

The glycolysis pathway, also known as the Embden-Meyerhof-Parnas Pathway or hexose diphosphate pathway, is the main pathway by which lactic acid bacteria metabolize monosaccharides (Wang et al., 2024). In the glycolytic pathway, expression of gene0518 (*pgi*), gene1188 (*pfk*A) and gene1189 (*pyk*) in mutant IMAU20551Y was lower than in the wild type (Fig. 5). Down-regulation of expression of these genes affected the ability of the mutant strain to metabolize glucose. This explains why glucose residues were still detected in milk fermented by the mutant strain after 14 d storage.

4. Discussion

S. thermophilus is the main bacterial strain used for yoghurt production, and can quickly convert lactose into lactic acid during milk fermentation, so that the pH of raw materials is rapidly reduced to promote curd, thus giving yoghurt excellent texture and taste (Erkus et al., 2014). *S. thermophilus* has been in long-term use in milk environment and as an adaptation to the single carbon source in dairy products, many genes related to carbon source transport, degradation and metabolism have been degraded into pseudogenes (Prajapati et al., 2013). The range of carbon sources available to *S. thermophilus* is narrow, mainly lactose and glucose, while use of galactose is strain specific (Gasser et al., 2022). Igoshi et al. (2017) found that galactose is the limiting factor leading to the Maillard reaction in ripe or overripe cheeses during storage. Therefore, we isolated a Gal $^{+}$ mutant strain IMAU20551 from wild type *S. thermophilus* and its galactose utilization rate was higher than that of the wild type.

Lactose enters the cell via lactose permease (LacS) and is hydrolyzed to glucose and galactose by the action of β-galactosidase (LacZ); and galactose is metabolized by the Leloir pathway (Zhang et al., 2019) (Fig. 7). At 0 h of fermentation, there was no significant difference in lactose content in fermented milk of the wild type and the mutant ($P > 0.05$) (Table 1). However, after fermentation and storage ended, the lactose content in the fermented milk of the mutant was lower than that of the wild type. The lactose that the mutant metabolizes more than the wild type will also be hydrolyzed into glucose for homolactic fermentation. As fermentation time increases, the galactose content in the fermented milk of the wild type continues to rise, while the galactose content in fermented milk of the mutant begins to decline after 4 h of fermentation. Galactose is first converted into galactose-1-phosphate under the action of galactokinase (GALK). Galactose-1-phosphate then undergoes a displacement reaction with UDP-glucose under the effect of galactose-1-phosphate uridyltransferase (GALT), generating UDP-galactose and glucose-1-phosphate. Glucose-1-phosphate is converted into glucose-6-phosphate by phosphoglucomutase (PGM) (Fig. 7). Therefore, this portion of galactose metabolized by the mutant is ultimately converted into glucose-6-phosphate, which enters the glycolysis pathway to participate in homolactic fermentation. Consequently, after fermentation and storage ended, the lactic acid content in

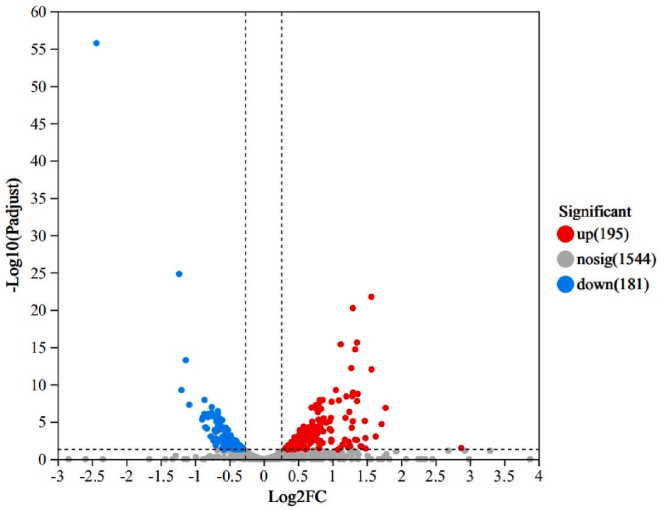


Fig. 4. Volcano map of differentially-expressed genes in *S. thermophilus* IMAU20551 and IMAU20551Y.

Table 3
KEGG, COG analysis and mutation type of mutated genes.

Gene ID	Gene Name	Mutation Type	Region Type	KEGG Definition	COG Description
gene0274	eno	snp	synonymous_variant	enolase	–
gene0923	gene0923	snp	stop_lost	–	magnesium and cobalt transport protein CorA
gene0958	potD	snp	missense_variant	spermidine/putrescine transport system substrate-binding protein	–
gene0617	metI	snp	missense_variant	D-methionine transport system permease protein	ABC transporter (permease)
gene0086	cwlO	del	inframe_deletion	peptidoglycan DL-endopeptidase CwlO	–
gene1653	gene1653	ins	frameshift_variant	putative transposase	–

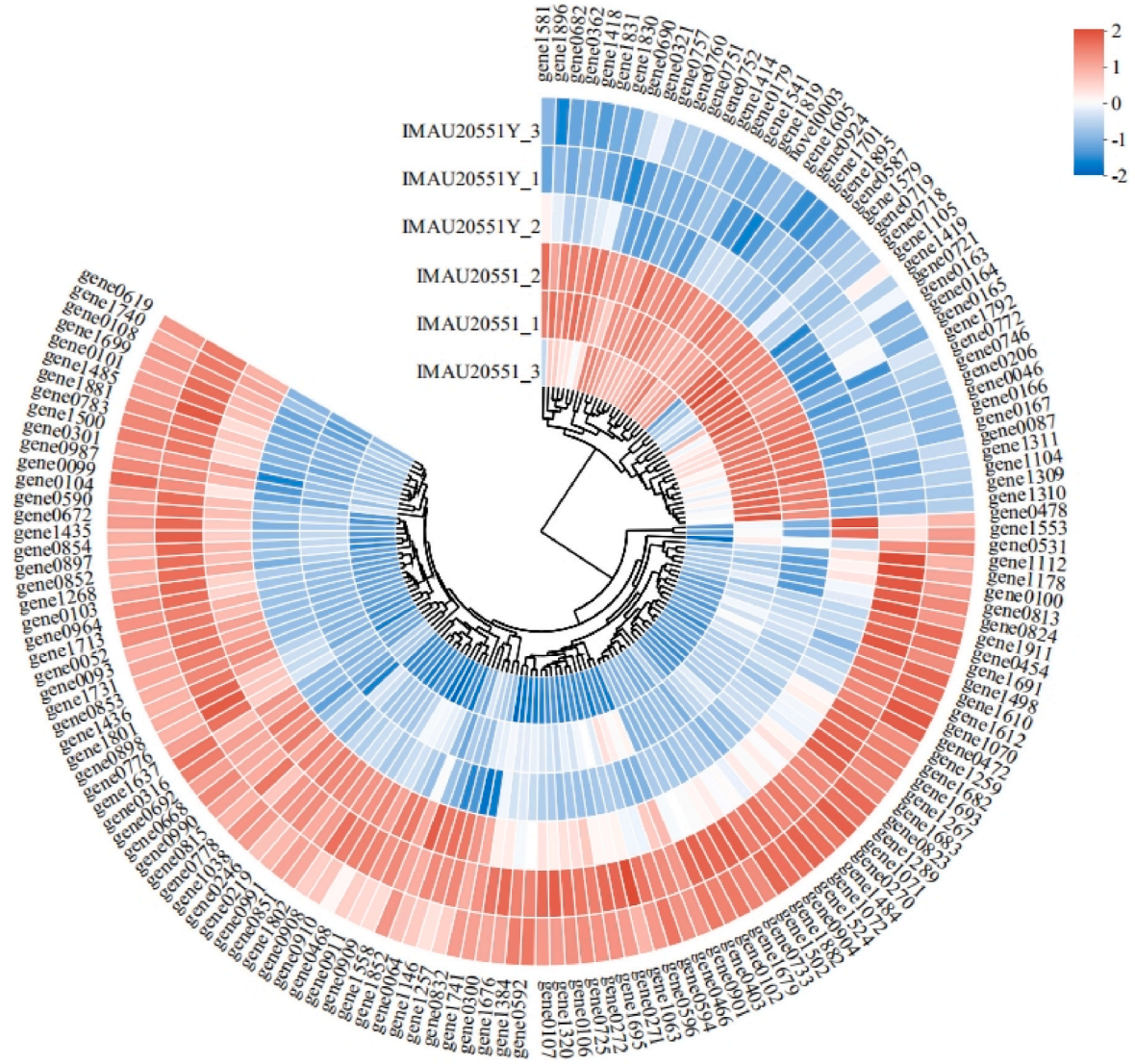


Fig. 5. Heatmap of differentially-expressed genes.

the fermented milk of the mutant was higher.

Genomic and transcriptomic analyses are necessary to reveal changes in the mutant strain at a molecular level (Okoye et al., 2022). Whole genome sequencing can obtain SNPs and InDels, which can be used to analyze genomic differences between strains and provide favorable conditions for studying gene function (Jia et al., 2023). Genomics results showed that the wild type IMAU20551 and the mutant strain IMAU20551Y were genetically very similar, but had a small number of SNPs and InDels variants in the CDS region. Wang et al. (2021) treated *Bacillus subtilis* JNFE0126 using ultraviolet radiation combined with ^{60}Co-γrays and found that mutant JNFE1126 had 28 SNPs and 15 Indels, and these mutant genes would cause phenotypic changes in the strain. Ge et al. (2022) showed that with variation in genome structure, many important genes that affect phenotype will also be missing or have an increase in copy number, which would also affect phenotype. We found mutations in genes encoding the ABC transporter and enolase (*metI*, *eno*) (Table 3). The ABC transporter transports a wide range of substrates important for normal physiological function and plays an important role in most organisms (Theodoulou and Kerr, 2015). Causey et al. (2004) found that the ABC transporter is related to membrane permeability and affects intracellular glucose transport, which in turn affects pyruvate metabolism. *eno* is a gene involved in pyruvate synthase, which can catalyze the formation of 2-phosphoglycerate into the high-energy compound phosphoenolpyruvate and control the rate of glycolysis (Cappello et al., 2017). Therefore, mutations in *metI* and *eno* genes may affect metabolic pathways of the mutant, and thus lead to phenotypic changes.

Transcriptomic results showed that expression of *pgi*, *pfk*A and *pyk* genes in the mutant was down-regulated (Fig. 5). The gene *pgi* encodes glucose-6-phosphate isomerase, which is located at branching points of the synthetic and catabolic pathways and mediates interconversion between glucose-6-phosphate and fructose-6-phosphate (Guo et al., 2017). Corominas et al. (1992) reported that decreased *pgi* activity would accumulate a large amount of glucose-6-phosphate and affect the rate of glycolytic metabolism. Therefore, the down-regulation of the *pgi* gene in the mutant leads to a decrease in synthesis and a reduction in activity of glucose-6-phosphate isomerase. This impedes conversion of glucose-6-phosphate to fructose-6-phosphate, slowing down the reaction rate of this crucial step in the glycolysis pathway. *pfk*A controls 6-phosphofructokinase irreversible phosphorylation of fructose-6-phosphate to fructose-1, 6-diphosphate (Li et al., 2023). In the mutant, down-regulation of *pfk*A gene leads to a decrease in synthesis of 6-phosphofructokinase, which reduces the rate of glycolysis. Since this reaction is one of the key rate-limiting steps in the glycolysis

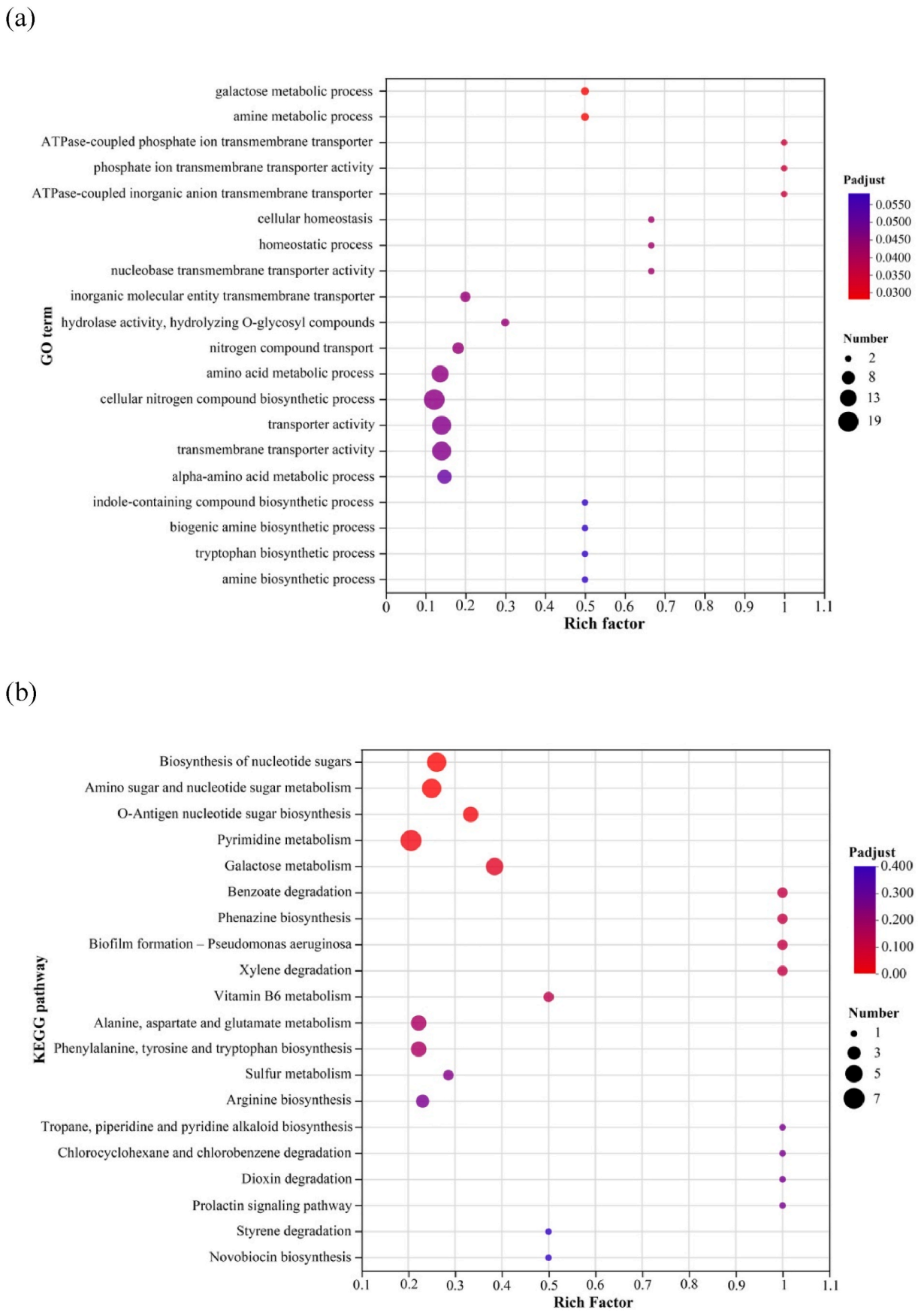


Fig. 6. GO enrichment analysis of differentially-expressed genes (a). KEGG enrichment analysis of differentially-expressed genes (b).

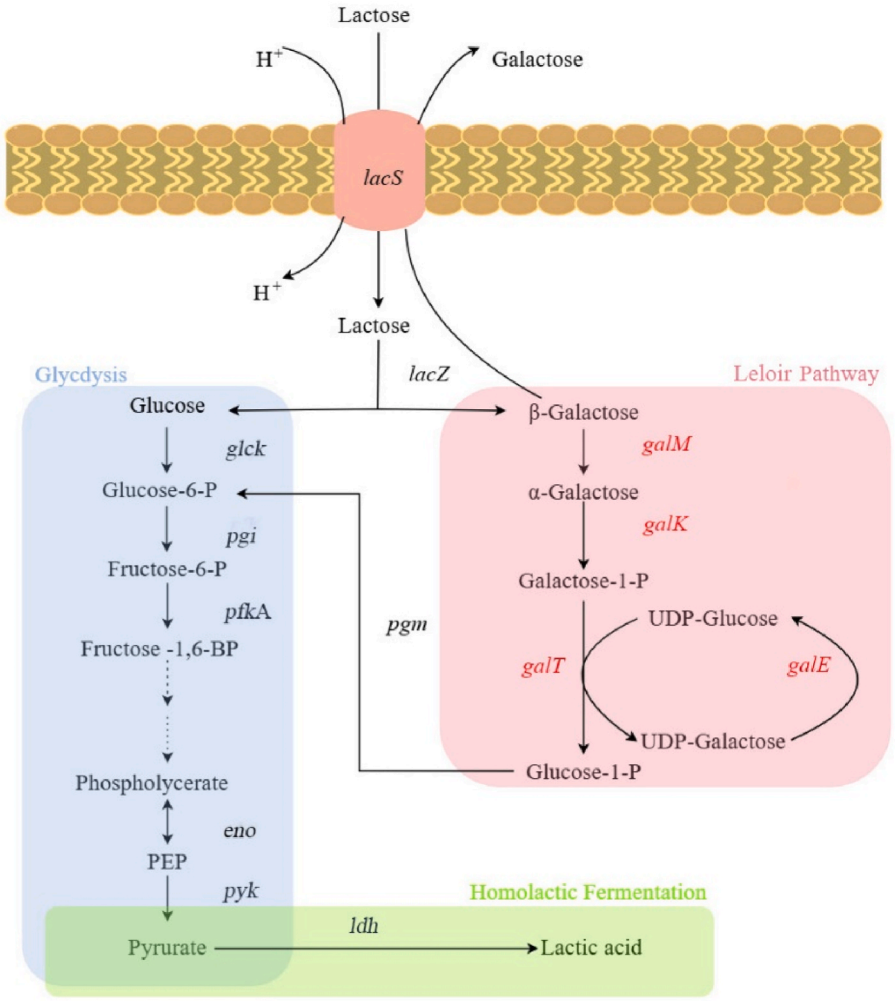


Fig. 7. Schematic representation of glycolysis/galactose metabolism in *S. thermophilus*.

process, the decrease in the reaction rate directly causes a reduction in the overall rate of glycolysis. The *pyk* gene encodes pyruvate kinase, which is involved in the last reaction step of glycolysis and is an important allosteric regulatory enzyme in the glycolysis pathway (Schormann et al., 2019). Pyruvate kinase is produced in all cells and tissues capable of glycolysis and its metabolites, ATP and pyruvate, are used in a variety of biosynthetic pathways (Zhang et al., 2012). Down-regulation of the *pyk* gene leads to an imbalance in allosteric regulation of pyruvate kinase, causing the rate of glycolysis to decline. Therefore, in the mutant, the down-regulated genes *pgi*, *pfk*A, and *pyk* inhibit the glycolysis metabolic process at different steps.

From transcriptome data it could be seen that the transcription levels of key enzyme genes in the mutant were significantly increased, e.g. the LacI family transcriptional regulator GalR and *galKTE* (Fig. 5). The Leloir metabolic pathway in *S. thermophilus* is related to the *galKTE* gene cluster (Vaillancourt et al., 2008). GalR is a transcriptional regulator of the Leloir pathway, encoded by genes upstream of *galKT*, and is mainly involved in regulating catabolic pathways such as lactose, galactose and ribose (McLean et al., 2020). Vaughan et al. (2001) confirmed that in *S. thermophilus*, GalR functions as a transcriptional activator of the gal operon. Therefore, *S. thermophilus* IMAU20551 and IMAU20551Y have different expression levels of gal genes co-transcribed on the same mRNA, which may be due to transcriptional regulation of GalR. When the intracellular galactose concentration is high, GalR binds to galactose, causing a conformational change in GalR. This change makes GalR dissociate from the operon sequence, removing the steric hindrance that blocks RNA polymerase from binding to the promoter. As a result, RNA polymerase can smoothly bind to the promoter region, initiating the transcription of the gal operon and effectively improving galactose metabolism (Semsey et al., 2006; Ajdić and Ferretti, 1998). De Vin et al. (2005) reported that increased transcription of the *galR-galK* gene region in Gal $^{+}$ mutants increased the rate of galactose metabolism. Therefore, up-regulation of GalR expression in the mutant can effectively activate expression of *galKTE* genes, enabling the cell to synthesize more key enzymes related to galactose metabolism and accelerating the intracellular galactose metabolism process. This explains why the mutant could metabolize more galactose during milk fermentation and storage compared with the wild type (Table 1).

Metabolism of galactose by Gal $^{+}$ *S. thermophilus* through the Leloir pathway is closely related to expression of the genes *galM*, *galK*, *galT* and *galE* (Anbukkarasi et al., 2014). Although we did not find mutations in the gene sequences associated with galactose metabolism through genome sequencing, transcriptome analysis showed that expression levels of *galM*, *galK*, *galT*, *galE* and LacI family transcriptional regulator GalR were up-regulated in the mutant. Ertl et al. (2022) reported that phenotypic evolution is often caused by changes in gene expression resulting from altered gene regulatory mechanisms. As a transcriptional activator of the *gal* operon, GalR can promote expression of *galKTE* genes related to galactose metabolism, thereby altering the phenotype of the mutant. The above factors could promote growth and metabolic activity of the mutant, and promote metabolism of galactose during milk fermentation and storage. In dairy production, *S. thermophilus* is commonly mixed with *Lactobacillus delbrueckii* subsp. *bulgaricus* (*L. delbrueckii* subsp. *bulgaricus*) for yogurt production. Therefore, it is

necessary to investigate the effect of strain combinations during fermentation on the metabolic pathways active. In addition, this study has led to further understanding of the mechanisms by which a Gal$^+$ mutant regulates metabolic pathways, including analysis of key metabolites, carbohydrates and genes/transcripts identified during milk fermentation and storage. Ultimately, the generated data will form the basis for selecting the ideal Gal$^+$ strains using genetic engineering.

5. Conclusion

In conclusion, this study verified differences in metabolic levels between the wild type *S. thermophilus* IMAU20551 and the mutant *S. thermophilus* IMAU20551Y by HPLC and found that the mutant could use more galactose. In addition, the internal mechanism for metabolic differences was defined using a combination of genomics and transcriptomics. Up-regulation of differentially-expressed genes *galM*, *galK*, *galT*, *galE* and LacI family transcriptional regulator GalR, was the fundamental cause of phenotypic change in the mutant. This study increased our understanding of the metabolic mechanisms of the Gal $^+$ *S. thermophilus* and provided a reference for genetic engineering modification of Gal $^+$ *S. thermophilus*, with a view to using it as a fermentation agent to improve the quality of dairy products, which is of great significance to the dairy industry.

CRediT authorship contribution statement

Jiahui Tai: Investigation, Methodology, Data curation, Writing – original draft, Writing – review & editing. **Haimin Hu:** Investigation, Methodology, Data curation. **Jinhui Liu:** Methodology, Data curation. **Wenhui Lu:** Methodology, Data curation. **Tong Dan:** Supervision, Data curation, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could influence the work reported in this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Beijing; No. 32072235), the Natural Science Foundation of Inner Mongolia (No. 2022MS03013), the Natural Science Foundation of Inner Mongolia (No. 2024LHMS03035).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crfs.2025.101017.

Data availability

The authors are unable or have chosen not to specify which data has been used.

References

Ajdić, D., Ferretti, J.J., 1998. Transcriptional regulation of the *Streptococcus mutans* gal operon by the GalR repressor. J. Bacteriol. 180 (21), 5727–5732. https://doi.org/10.1128/jb.180.21.5727-5732.1998.

Anbukkarasi, K., Nanda, D.K., UmaMaheswari, T., Hemalatha, T., Singh, P., Singh, R., 2014. Assessment of expression of Leloir pathway genes in wild-type galactose-fermenting *Streptococcus thermophilus* by real-time PCR. Eur Food Res Technol 239, 895–903. https://doi.org/10.1007/s00217-014-2286-9.

Benateya, A., Bracquart, P., Linden, G., 1991. Galactose-fermenting mutants of *Streptococcus thermophilus*. Can. J. Microbiol. 37 (2), 136–140. https://doi.org/10.1139/m91-020.

Cappello, P., Principe, M., Bulfamante, S., Novelli, F., 2017. Alpha-Enolase (*ENO1*), a potential target in novel immunotherapies. Front Biosci-Landmrk 22 (5), 944–959. https://doi.org/10.2741/4526.

Causey, T.B., Shanmugam, K.T., Yomano, L.P., Ingram, L.O., 2004. Engineering *Escherichia coli* for efficient conversion of glucose to pyruvate. Proc. Natl. Acad. Sci. U.S.A. 101 (8), 2235–2240. https://doi.org/10.1073/pnas.0308171100.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6 (2), 80–92. https://doi.org/10.4161/fly.19695.

Conte, F., van Buuringen, N., Voermans, N.C., Lefeber, D.J., 2021. Galactose in human metabolism, glycosylation and congenital metabolic diseases: time for a closer look. Biochim. Biophys. Acta Gen. Subj. 1865 (8), 129898. https://doi.org/10.1016/j.bbagen.2021.129898.

Corominas, J., Clotet, J., Fernández-Bañares, I., Boles, E., Zimmmermann, F.K., Guinovart, J.J., Ariño, J., 1992. Glycogen metabolism in a *Saccharomyces cerevisiae* phosphoglucose isomerase (*pgi1*) disruption mutant. FEBS Lett. 310 (2), 182–186. https://doi.org/10.1016/0014-5793(92)81325-G.

Cui, Y., Xu, T., Qu, X., Hu, T., Jiang, X., Zhao, C., 2016. New insights into various production characteristics of *Streptococcus thermophilus* strains. Int. J. Mol. Sci. 17 (10), 1701. https://doi.org/10.3390/ijms17101701.

De Vin, F., Rådström, P., Herman, L., De Vuyst, L., 2005. Molecular and biochemical analysis of the galactose phenotype of dairy *Streptococcus thermophilus* strains reveals four different fermentation profiles. Appl. Environ. Microbiol. 71 (7), 3659–3667. https://doi.org/10.1128/AEM.71.7.3659-3667.2005.

Erkus, O., Okuklu, B., Yenidunya, A.F., Harsa, S., 2014. High genetic and phenotypic variability of *Streptococcus thermophilus* strains isolated from artisanal Yuruk yoghurts. LWT-Food Sci. Technol 58 (2), 348–354. https://doi.org/10.1016/j.lwt.2013.03.007.

Ertl, H.A., Hill, M.S., Wittkopp, P.J., 2022. Differential Grainy head binding correlates with variation in chromatin structure and gene expression in Drosophila melanogaster. BMC Genom. 23 (1), 854. https://doi.org/10.1186/s12864-022-09082-7.

Gasser, C., Garault, P., Chervaux, C., Monnet, V., Faurie, J.M., Rul, F., 2022. Co-utilization of saccharides in mixtures: Moving toward a new understanding of carbon metabolism in *Streptococcus thermophilus*. Food Microbiol. 107, 104080. https://doi.org/10.1016/j.fm.2022.104080.

Ge, H., Li, G., Wan, S., Zhao, A., Huang, Y., Ma, R., Zhang, R., Song, Y., Sha, G., 2022. Whole genome re-sequencing and transcriptome reveal an alteration in hormone signal transduction in a more-branching mutant of apple. Gene 818, 146214. https://doi.org/10.1016/j.gene.2022.146214.

Guo, W., Zou, L., Ji, Z., Cai, L., Chen, G., 2017. Glucose 6-phosphate isomerase (*Pgi*) is required for extracellular polysaccharide biosynthesis, DSF signals production and full virulence of *Xanthomonas oryzae* pv. *oryzicola* in rice. Physiol. Mol. Plant Pathol. 100, 209–219. https://doi.org/10.1016/j.pmpp.2017.10.010.

Harper, M., Lee, C.J., 2012. Genome-wide analysis of mutagenesis bias and context sensitivity of N-methyl-N-nitro-N-nitrosoguanidine (NTG). Mutat. Res. 731 (1–2), 64–67. https://doi.org/10.1016/i.mrfmmm.2011.10.011.

Hu, H.F., Zhou, H.Y., Cheng, G.P., Xue, Y.P., Wang, Y.S., Zheng, Y.G., 2020. Improvement of R-2-(4-hydroxyphenoxy) propionic acid biosynthesis of *Beauveria bassiana* by combined mutagenesis. Biotechnol. Appl. Biochem. 67 (3), 343–353. https://doi.org/10.1002/bab.1872.

Hu, H., Nie, J., Wu, N., Dan, T., 2024. Selection of a galactose-positive mutant strain of *Streptococcus thermophilus* and its optimized production as a high-vitality starter culture. J. Dairy Sci. 107 (9), 6558–6575. https://doi.org/10.3168/jds.2023-24550.

Igoshi, A., Sato, Y., Kameyama, K., Murata, M., 2017. Galactose is the limiting factor for the browning or discoloration of cheese during storage. J. Nutr. Sci. Vitaminol. 63 (6), 412–418. https://doi.org/10.3177/jnsv.63.412.

Iskandar, C.F., Cailliez-Grimal, C., Borges, F., Revol-Junelles, A.M., 2019. Review of lactose and galactose metabolism in Lactic Acid Bacteria dedicated to expert genomic annotation. Trends Food Sci. Technol. 88, 121–132. https://doi.org/10.1016/j.tifs.2019.03.020.

Jia, Y., Shen, T., Wen, Z., Chen, J., Liu, Q., 2023. Combining transcriptome and whole genome Re-sequencing to screen disease resistance genes for wheat dwarf bunt. Int. J. Mol. Sci. 24 (24), 17356. https://doi.org/10.3390/ijms242417356.

Johnson, M.E., Olson, N.F., 1985. Nonenzymatic browning of Mozzarella cheese. J. Dairy Sci. 68 (12), 3143–3147. https://doi.org/10.3168/jds.S0022-0302(85)81219-4.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., Petrov, A.I., 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 46 (D1), D335–D342. https://doi.org/10.1093/nar/gkx1038.

Kosugi, S., Terao, C., 2024. Comparative evaluation of SNVs, indels, and structural variations detected with short-and long-read sequencing data. Hum Genome Var 11 (1), 18. https://doi.org/10.1038/s41439-024-00276-x.

Li, E., Li, S., Liu, F., Li, Q., Pang, D., Wang, H., Liao, S., Zou, Y., 2023. Analysis of *Akkermansia muciniphila* in mulberry galacto-oligosaccharide medium via comparative transcriptomics. Foods 12 (3), 440. https://doi.org/10.3390/foods12030440.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1 (1), 2047. https://doi.org/10.1186/2047-217X-1-18, 217X.

Mathlin, J., Le Pera, L., Colombo, T., 2020. A census and categorization method of epitranscriptomic marks. Int. J. Mol. Sci. 21 (13), 4684. https://doi.org/10.3390/ijms21134684.

McLean, K.T., Tikhomirova, A., Brazel, E.B., Legendre, S., Haasbroek, G., Minhas, V., Paton, J.C., Trappetti, C., 2020. Site-specific mutations of GalR affect galactose metabolism in *Streptococcus pneumoniae*. J. Bacteriol. 203 (1), 10–1128. https://doi.org/10.1128/jb.00180-20.

Okoye, C.O., Dong, K., Wang, Y., Gao, L., Li, X., Wu, Y., Jiang, J., 2022. Comparative genomics reveals the organic acid biosynthesis metabolic pathways among five lactic acid bacterial species isolated from fermented vegetables. N Biotechnol 70, 73–83. https://doi.org/10.1016/j.nbt.2022.05.001.

Potts, A.S., Hunter, M.D., 2021. Unraveling the roles of genotype and environment in the expression of plant defense phenotypes. Ecol. Evol. 11 (13), 8542–8561. https://doi.org/10.1002/ece3.7639.

Prajapati, J.B., Nathani, N.M., Patel, A.K., Senan, S., Joshi, C.G., 2013. Genomic analysis of dairy starter culture *Streptococcus thermophilus* MTCC 5461. J. Microbiol. Biotechnol. 23 (4), 459–466. https://doi.org/10.4014/jmb.1210.10030.

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R.P., Banday, S., Mishra, A.K., Das, G., Malonia, S.K., 2023. Next-generation sequencing technology: current trends and advancements. Biology 12 (7), 997. https://doi.org/10.3390/biology12070997.

Schormann, N., Hayden, K.L., Lee, P., Banerjee, S., Chattopadhyay, D., 2019. An overview of structure, function, and regulation of pyruvate kinases. Protein Sci. 28 (10), 1771–1784. https://doi.org/10.1002/pro.3691.

Semsey, S., Virnik, K., Adhya, S., 2006. Three-stage regulation of the amphibolic gal operon: from repressosome to GalR-free DNA. J. Mol. Biol. 358 (2), 355–363. https://doi.org/10.1016/j.jmb.2006.02.022.

Shahwar, D., Ahn, N., Kim, D., Ahn, W., Park, Y., 2023. Mutagenesis-based plant breeding approaches and genome engineering: a review focused on tomato. Mutat. Res. Rev. Mutat. Res., 108473 https://doi.org/10.1016/j.mrrev.2023.108473.

Stothard, P., Grant, J.R., Van Domselaar, G., 2019. Visualizing and comparing circular genomes using the CGView family of tools. Brief Bioinform 20 (4), 1576–1582. https://doi.org/10.1093/bib/bbx081.

Sun, L., Lu, Z., Li, J., Sun, F., Huang, R., 2018. Comparative genomics and transcriptome analysis of *Lactobacillus rhamnosus* ATCC 11443 and the mutant strain SCT-10-10-60 with enhanced L-lactic acid production capacity. Mol Genet Genomics 293, 265–276. https://doi.org/10.1007/s00438-017-1379-0.

Theodoulou, F.L., Kerr, I.D., 2015. ABC transporter research: going strong 40 years on. Biochem. Soc. Trans. 43 (5), 1033–1040. https://doi.org/10.1042/BST20150139.

Vaillancourt, K., Bédard, N., Bart, C., Tessier, M., Robitaille, G., Turgeon, N., Frenette, M., Moineau, S., Vadeboncoeur, C., 2008. Role of *galK* and *galM* in galactose metabolism by *Streptococcus thermophilus*. Appl. Environ. Microbiol. 74 (4), 1264–1267. https://doi.org/10.1128/AEM.01585-07.

Vaughan, E.E., van den Bogaard, P.T., Catzeddu, P., Kuipers, O.P., de Vos, W.M., 2001. Activation of silent gal genes in the lac-gal regulon of *Streptococcus thermophilus*. J. Bacteriol. 183 (4), 1184–1194. https://doi.org/10.1128/jb.183.4.1184-1194.2001.

Wang, Y., Peng, Q., Liu, Y., Wu, N., He, Y., Cui, X., Dan, T., 2024. Genomic and transcriptomic analysis of genes involved in exopolysaccharide biosynthesis by *Streptococcus thermophilus* IMAU20561 grown on different sources of nitrogen. Front. Microbiol. 14, 1328824. https://doi.org/10.3389/fmicb.2023.1328824.

Wang, Y., Wang, J., Zhang, X., Tong, Y., Yang, R., 2021. Genomic and transcriptomic analysis of *Bacillus subtilis* JNFE1126 with higher nattokinase production through ultraviolet combined ^{60}Co-γ ray mutagenesis. Lwt-Food Sci Technol 147, 111652. https://doi.org/10.1016/j.lwt.2021.111652.

Wu, Q., Cheung, C.K., Shah, N.P., 2015. Towards galactose accumulation in dairy foods fermented by conventional starter cultures: challenges and strategies. Trends Food Sci. Technol. 41 (1), 24–36. https://doi.org/10.1016/j.tifs.2014.08.010.

Xu, Z.S., Liang, Y., Kong, J., Zhang, S.S., Liu, X.L., Wang, T., 2022. A food-grade vector for *Streptococcus thermophilus* based on the α-complementation of β-galactosidase. J. Dairy Sci. 105 (7), 5641–5653. https://doi.org/10.3168/jds.2021-21699.

Yang, Y., Okoye, C.O., Xiang, J., Huang, W., Liu, Y., He, R., Huang, G., Ma, H., 2024. Genome-wide transcriptome analyses reveal changes in glutathione-overproducing yeast obtained by ARTP mutagenesis for rice wine brewing. Food Biosci. 60, 104388. https://doi.org/10.1016/j.fbio.2024.104388.

Zhang, X., Lai, L., Xu, G., Zhang, X., Shi, J., Koffas, M.A., Xu, Z., 2019. Rewiring the central metabolic pathway for high-yield L-serine production in *Corynebacterium glutamicum* by using glucose. Biotechnol. J. 14 (6), 1800497. https://doi.org/10.1002/biot.201800497.

Zhang, Y., Xiao, W., Luo, L., Pang, J., Rong, W., He, C., 2012. Downregulation of *OsPK1*, a cytosolic pyruvate kinase, by T-DNA insertion causes dwarfism and panicle enclosure in rice. Planta 235 (1), 25–38. https://doi.org/10.1007/s00425-011-1471-3.