Received: 30 March 2019 | Revised: 19 July 2019 | Accepted: 25 July 2019

DOI: 10.1002/mbo3.922

ORIGINAL ARTICLE

MicrobiologyOpen Open Access WILEY

The complete genome sequence of the thermophilic bacterium *Laceyella sacchari* FBKL4.010 reveals the basis for tetramethylpyrazine biosynthesis in Moutai-flavor Daqu

Dounan Li[1,2] | Wei Huang[2,3] | Chunxiao Wang[2,3] | Shuyi Qiu[1,2,3]

[1]College of Life Sciences, Guizhou University, Guiyang, China

[2]The Key Laboratory of Fermentation Engineering and Biological Pharmacy of Guizhou Province, Guizhou University, Guiyang, China

[3]School of Liquor-Making and Food Engineering, Guizhou University, Guiyang, China

Correspondence
Shuyi Qiu, College of Life Sciences, Guizhou University, Guiyang, Guizhou, China.
Email: syqiu@gzu.edu.cn

Funding information
Special Project in Guizhou Province, Grant/Award Number: GZ [2015]6012; Industrial Research Project in Guizhou Province, Grant/Award Number: GZ [2014]3012

Abstract

The genus *Laceyella* consists of a thermophilic filamentous bacteria. The pure isolate of *Laceyella sacchari* FBKL4.010 was isolated from Moutai-flavor Daqu, Guizhou Province, China. In this study, the whole genome was sequenced and analyzed. The complete genome consists of one 3,374,379-bp circular chromosome with 3,145 coding sequences (CDSs), seven clustered regularly interspaced short palindromic repeat (CRISPR) regions of 12 CRISPRs. Moreover, we identified that the genome contains genes encoding key enzymes such as proteases, peptidases, and acetolactate synthase (ALS) of the tetramethylpyrazine metabolic pathway. Metabolic pathways relevant to tetramethylpyrazine synthesis were also reconstructed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database. Annotation and syntenic analyses using antiSMASH 4.0 also revealed the presence of two gene clusters in this strain that differ from known tetramethylpyrazine synthesis clusters, with one encoding amino acid dehydrogenase (ADH) and the other encoding transaminase in tetramethylpyrazine metabolism. The results of this study provide flavor and genomic references for further research on the flavor-producing functions of strain FBKL4.010 in the Moutai liquor-making process.

KEYWORDS

antiSMASH analysis, complete genome, *Laceyella sacchari*, liquor-making, Moutai-flavor Daqu, tetramethylpyrazine

1 | INTRODUCTION

Chinese Moutai liquor has been made for 800 years and is distilled from the product of fermentation using a wild microbial starter, Moutai-flavor Daqu. In contrast to other Daqus, this Daqu uses only wheat as raw material and is an important starter for the whole fermentation process (Wang, Shi, & Gong, 2008). In general, the maximum temperature while making Moutai-flavor can reach 65°C. This temperature creates good growth or sporulation conditions for many thermophilic microbes that are believed to be important in the formation of the soy sauce flavor during the liquor-making process (Li, Lian, Ding, Nie, & Zhang, 2014; Wang et al., 2017; Zhang et al., 2012). Among these microbes, *Bacillus* sp. such as *Bacillus subtilis* and *Bacillus licheniformis* have been considered as a source of the soy sauce flavor in Moutai liquor. However, the roles of other thermophilic microbes in the formation of the soy sauce flavor have not been confirmed (Li & Qiu, 2017).

Laceyella sacchari is a chemoheterotrophic, heat-resistant, and gram-positive microbe in the family *Thermoactinomycetaceae*, genus *Laceyella*. Members of the genus *Laceyella* originally belonged to the

genus *Thermoactinomyces*, which was first discovered and described as a single species by Tsiklinsky in 1899 (Yoon, Kim, Shin, & Park, 2005). After the use of new polyphasic taxonomy methods, the genus *Laceyella* was divided from the genus *Thermoactinomyces* and described as a new genus (Matsuo et al., 2006; Yoon et al., 2005). *Laceyella sacchari* often inhabits high-temperature environments (Xian, Ming, & Li, 2015) and has great potential for application in the modern food industry (Hanphakphoom, Maneewong, Sukkhum, Tokuyama, & Kitpreechavanich, 2014; Shukla & Singh, 2015).

According to the whole-genome analysis of the phylogenetically related strain, *Thermoactinomyces daqus* H-18, which was isolated from high-temperature Daqu (approximately 3.44Mb), 60 coding sequences (CDSs) of peptidases and amino acid dehydrogenases (ADHs) were identified (Yao et al., 2014). In the tetramethylpyrazine metabolic pathway, peptidases degrade proteins into amino acids and ADH further turns amino acids into ammonium (an important flavor precursor) (Wu & Xu, 2014; Zhu, Xu, & Fan, 2010). This result revealed that some thermophilic microbes in the family *Thermoactinomycetaceae* in Daqu may play a role similar to that of *Bacillus* sp., which is relevant to the formation of soy sauce flavor. In a recent study by our research group, the family *Thermoactinomycetaceae* was identified as a novel dominant microbial group (accounting for 34.40% of the total microbial biomass) in Moutai-flavor Daqu through metagenome sequencing (Wang, Ban, Zhou, Hu, & Qiu, 2016). A pure isolate designated FBKL4.010 was obtained from Moutai-flavor Daqu and identified as *L. sacchari* using gene phylogenetic analysis based on the 16S rRNA sequence. Thus, the aims of this study were (a) to explore the main function of the strain FBKL4.010 under a simulated environment of Moutai-flavor Daqu fermentation and (b) to confirm key genes involved in flavor formation on the genome of FBKL4.010.

2 | MATERIALS AND METHODS

2.1 | Target strain

The strain FBKL4.010 was isolated from Moutai-flavor Daqu and identified as *L. sacchari* in our previous study.

2.2 | GC-MS analysis

In the solid-state fermentation experiment, we used pure wheat as a solid-state medium to simulate the fermentation of Moutai-flavor Daqu. The wheat was moistened in 60% boiling water for 5–6 hr and saccharified in a 60°C water bath for 3–4 hr. Finally, we sterilized the medium at 115°C for 30 min. FBKL4.010 was cultured in sterilized wheat medium at its optimum growth temperature (45°C). After 5 days of cultivation in 250-ml bottles under aerobic, standing conditions, we selected fermentative material as the subject for gas chromatography–mass spectrometry (GC-MS) analysis. Fermentation samples (5 g per sample) were weighed into 20-ml vials sealed with Teflon-coated septa and incubated in a 60°C water bath for 15 min prior to a 30-min solid-phase microextraction (SPME) extraction. The SPME fiber assembly, 50/30 μm DVB/CAR/PDMS, stableflex, manual holder was obtained from Supelco. Samples of the medium alone were used as the blank group. Substances were separated with a ZB-5MSI column (30 m × 0.25 mm × 0.25 μm). The injection temperature was 250°C (splitless injection mode), and high-purity He (99.999%) was used as the carrier gas with a 1.0 ml/min flow rate. The column pressure was set at 7.62 psi, and the solvent delay time was 1 min. The temperature program was as follows: the oven initial temperature was maintained at 40°C for 6 min, increased to 135°C at 4°C/min, and maintained for 2 min, and then increased to 180°C at 8°C/min and maintained for 5 min. The ion source temperature was 250°C; the MS quad temperature was 150°C; the ionization voltage was 70 eV; and the emission current was 34.6 μA. The scan range was 20–450 amu.

2.3 | DNA extraction and genomic sequencing

Genomic DNA of FBKL4.010 was purified using the DNeasy Blood & Tissue Kit (QIAGEN®). About 155 μg DNA was obtained taken up in the water, and 80.7 μg was used for sequencing (the concentration of DNA was 180.54 ng/μl, with an $OD_{260/280}$ value of 2.06). Whole-genome sequencing was performed on the Illumina MiSeq-Dx Sequencer using a PE400 library (400-bp inserts) in the sequencing mode of paired-end 2 × 251 bp and PacBio RS II Sequencer using a S20K library (20-kbp inserts) with P6-C4 chemistry in the standard sequencing mode. Raw reads from the Illumina MiSeq platform were examined and subjected to quality control using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Low-quality reads were removed using SOAPec v2.0 (Luo et al., 2012), and high-quality reads were obtained based on the kmer frequency (value = 17). Subsequently, the FBKL4.010 genome was assembled using SPAdes v3.9.0 (Bankevich et al., 2012). The data from the PacBio platform were assembled into a number of scaffolds and contigs using Canu v1.4 (Koren et al., 2017). Finally, we generated a mixed assembly using SPAdes v3.9.0 (Bankevich et al., 2012) and a multicollinearity analysis using MUMmer v3.1 (Kurtz et al., 2004) to close the gaps among the contigs.

2.4 | Genomic functional annotation and discovery of tetramethylpyrazine-relevant genes

The protein-coding genes were predicted using Glimmer 3.02 (Delcher, Harmon, Kasif, White, & Salzberg, 1999) and the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline with GeneMarkS+ v4.3 (Tatusova et al., 2016). In addition, annotation was further refined with blast searches to the NCBI Reference Sequence database (RefSeq) (Release on March 20, 2019) and the nonredundant protein sequence database (NCBI-nr) (Pruitt, Tatusova, Klimke, & Maglott, 2009) using blast+ v2.8.1 (Mount, 2007). The functional criteria included an E-value of less than 1e-6, the sequence identity more than 30%, and the minimal alignment length percentage of larger than 70%. Finally, the annotated information of the best hit was assigned to the corresponding

protein-coding gene. The discovery of tetramethylpyrazine-relevant genes and reconstruction of relevant metabolic pathways was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY and Mapper databases (Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016). The prediction of gene clusters and syntenic analyses were performed using antiSMASH 4.0, bacterial version (Tilmann et al., 2015). The circular graph of the genome was generated using CGView (Copyright 2004) (Stothard & Wishart, 2005).

3 | RESULTS AND DISCUSSION

3.1 | Flavor component characteristics of FBKL4.010

According to the results of the GC-MS analysis, we found that the strain FBKL4.010 produces a flavor precursor, acetoin, and various pyrazines, including 2,5-dimethyl pyrazine, 2,3,5-trimethylpyrazine, tetramethylpyrazine, and 2-ethyl-3,5,6-trimethylpyrazine, under a solid-state system using pure wheat (Figure 1). Notably, the most abundant flavor component was tetramethylpyrazine (red frame in Figure 1), which accounted for 52.68% of the total flavor component content. In previous GC-MS analyses of *B. subtilis* MTDB-03, *B. licheniformis* FBKL1.0199 and FBKL1.0201 isolated from Moutai-flavor Daqu, similar flavor components were also identified, with the tetramethylpyrazine content exceeded 40.00% of the total flavor component for each strain (Wang et al., 2017; Yang et al., 2011). In the previous study, no strains in the genus *Laceyella* with the ability to produce tetramethylpyrazine have been described. Therefore, the identification of a strain in the genus *Laceyella* that plays a role in the formation of soy sauce flavor, similar to that of *Bacillus* sp. under the simulated environment of Moutai-flavor Daqu fermentation, is notable.

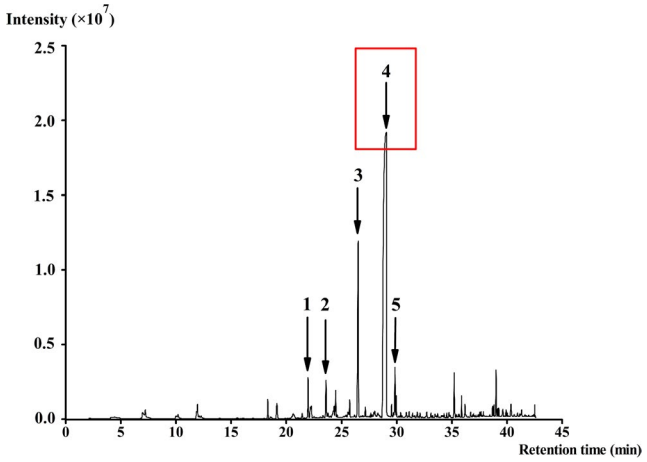

FIGURE 1 The chromatogram of flavor components under solid-state fermentation conditions from the strain FBKL4.010. Number 1 represents the precursor acetoin; numbers 2, 3, 4, and 5 represent 2,5-dimethyl pyrazine, 2,3,5-trimethylpyrazine, tetramethylpyrazine, and 2-ethyl-3,5,6-trimethylpyrazine, respectively

3.2 | Genome characteristics

Further research was conducted to identify key genes relevant to tetramethylpyrazine metabolism using complete genome analysis to study the tetramethylpyrazine metabolic pathway in FBKL4.010 under solid fermentation conditions. The results showed that the complete genome sequence of strain FBKL4.010 consisted of a single, circular 3,374,379-bp chromosome (Figure 2) based on the mixed assembly and multicollinearity analysis. The GenBank accession no. is CP025943.1. The average GC content of the whole sequence is 49.19%. The analysis with the Prokaryotic Genome Annotation Pipeline (GeneMarkS+ v4.3) (Tatusova et al., 2016) resulted in the annotation of 3,145 CDSs, 12 clustered regularly interspaced short palindromic repeats (CRISPRs) (distributed in 7 regions) and 135 RNAs, including 97 tRNAs, 34 rRNAs, and 4 other ncRNAs in the whole-genome sequence (Table 1). These genome features are similar to other members in the genus *Laceyella* such as *L. sacchari* 1–1 (GenBank WGS accession number is ASZU00000000.1), which has a genome size of 3.32 Mbp and 48.90% G + C content (ANI value with FBKL4.010 was 97.81%) (Kaur, Arora, Kumar, & Mayilraj, 2014), and *Laceyella sediminis* RHA1 (GenBank WGS accession number is NZ_PVTZ00000000), which has a genome size of 3.38 Mbp and 48.90% G + C content, 75 tRNAs, 39 rRNAs, and 4 ncRNAs (ANI value with FBKL4.010 was 95.57%) (Whitman et al., 2015). FBKL4.010 can be classified into *L. sacchari* due to the high ANI values observed above the 95% threshold.

3.3 | Tetramethylpyrazine-relevant genes and reconstruction of relevant metabolic pathways

Tetramethylpyrazine, which often present in Chinese liquor-making systems, is considered to be a major source of soy sauce flavor (Xu, Wu, Fan, & Zhu, 2011) and is a component of Moutai-flavor liquor (Huo et al., 2017). Tetramethylpyrazine is generated by the Maillard reaction during the Daqu-making process and is transferred to enter the liquor by distillation (Wu, 2007). With the rapid development of the Chinese liquor industry, research on tetramethylpyrazine has become a major focus (Zhu et al., 2010). In general, the metabolic pathways of acetoin and ammonium (NH_3) are two important aspects of microbes producing tetramethylpyrazine in the solid-state system of Chinese liquor-making. The acetoin biosynthetic pathway of in bacteria such as *B. subtilis* has been well studied (Xiao, Hou, Xin, Xi, & Zhao, 2014; Xiao, Liu, Qin, & Xu, 2007; Xu et al., 2011; Zhu, Xu, & Fan, 2009). The formation of NH_3 under solid substrate conditions has also been analyzed in recent research (Wu & Xu, 2014). All of the metabolic pathways and key enzymes relevant to tetramethylpyrazine production in *B. subtilis* under solid-state fermentation conditions are shown in Figure 3. However, not all genes encoding key enzymes in the pathway were identified in the strain in this study, as indicated in Figure 3 with different colors. However, annotations for two metabolic pathways, the acetoin and NH_3 metabolic pathways, were identified in *L. sacchari* FBKL4.010.

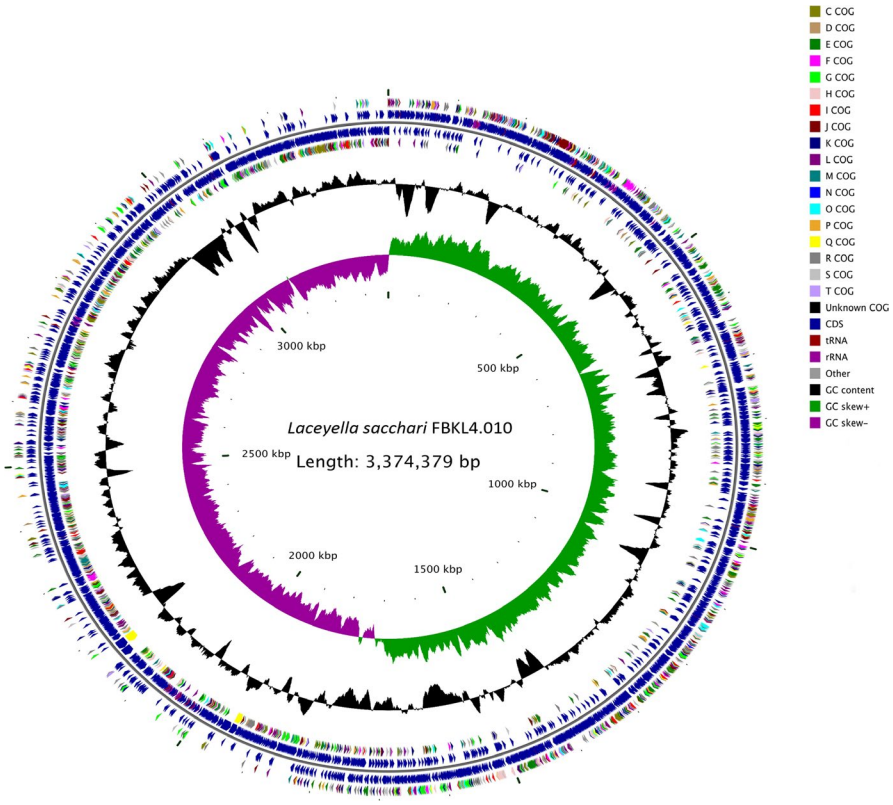

FIGURE 2 Circular graph of *Laceyella sacchari* FBKL4.010. The innermost circle represents the GC Skew +(green) and GC Skew - (purple), followed by the second circle representing the GC content (black); the outside rings depict the position of RNAs or clusters of orthologous group (COG) annotations of various structures such as CDSs in the genome of the strain

TABLE 1 Genome features of strain FBKL4.010

Features	Chromosome
Genome size (bp)	3,374,379
G + C content	49.19%
Average genome coverage	202× (or fold)
Plasmid	0
CDS (total)	3,199
CDS (coding)	3,145
tRNAs	97
rRNAs	34
ncRNAs	4
CRISPRs	12
Putative genes	3,334
Prophage (incomplete)	2
GenBank accession no.	CP025943.1

Gene functional annotations of *L. sacchari* FBKL4.010 allowed us to identify genes encoding key enzymes that are relevant to acetoin metabolism, such as the *ilvB* and *ilvN* genes, which encode the large and small subunits of acetolactate synthase (ALS), respectively (GenBank accession no. of protein sequences: AUS09165.1 and AUS09164.1). The bacterial ALS holoenzyme is a tetramer of two identical large subunits and two identical small subunits encoded by two genes of the large subunit and small subunit, respectively (Hershey, Schwartz, Gale, & Abell, 1999). This study is the first time that these genes have been identified in *Laceyella sacchari*, and they are similar to those of *L. sediminis* (Chen et al., 2012), and *Thermoactinomycetaceae* with 98% (large SU), 95% (small SU) amino acid sequence identities (GenBank accession no. of protein sequences: WP_106343376.1 and WP_022739119.1). However, in our study, key genes encoding acetolactate decarboxylase (ALDC) and 2,3-butanediol dehydrogenase (BDH) could not be identified in the genome of strain FBKL4.010 (green circles in Figure 3). It was supposed that an unknown alternative route of acetoin biosynthesis exists in this strain.

Functional annotations also pinpointed genes in NH_3 metabolic pathways that encoded proteases, peptidases, and carboxypeptidases. It has been shown that these enzymes are often involved in protein and peptide degradation and in the degradation of proteins into free amino acids (the detailed information is shown in Table 2). Some of these enzymes have good thermostability and thermophilic characteristics (Arndt et al., 2002), indicating that they may play a role in the high-temperature Daqu-making process, such as Moutai-flavor Daqu.

In addition, the genome of strain FBKL4.010 contains genes that encoded ADHs, such as alanine, glutamate, leucine, and proline dehydrogenases (GenBank accession no. of protein sequences: AUS08132.1, AUS08496.1, AUS08677.1, AUS08555.1, AUS09758.1, and AUS08130.1). In a previous study, ADH and aminotransferase were shown to play a role in enzymatic transamination and transdeamination. Free amino acids are transformed into NH_3 via catalysis by these enzymes (Wu & Xu, 2014). The identification of all genes

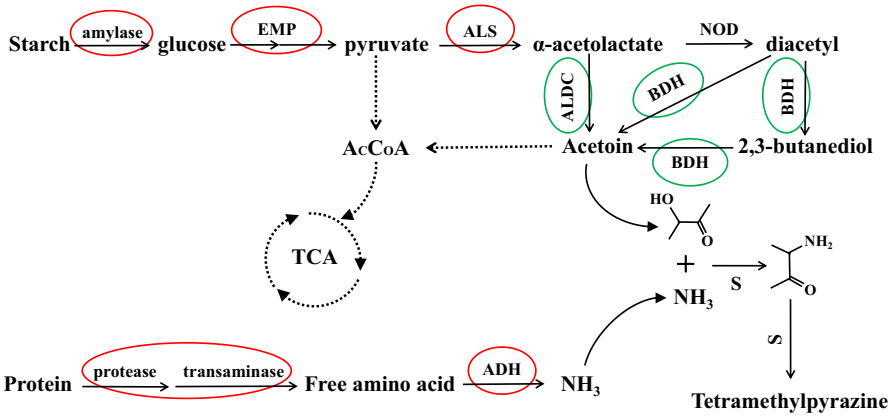

FIGURE 3 The *Bacillus subtilis* tetramethylpyrazine metabolic pathway under solid-state fermentation conditions in Chinese liquor-making (Wu & Xu, 2014). Bold arrows with solid lines are the major reactions taking place during tetramethylpyrazine formation. Arrows with dotted lines indicate branched metabolic pathways. Red circles and green circles indicate the confirmed and unconfirmed genes encoding key enzymes in this study, respectively. ALS, acetolactate synthase; ALDC, α-acetolactate decarboxylase; NOD, the enzyme catalytic oxidative decarboxylation; BDH, 2,3-butanediol dehydrogenase; ADH, amino acid dehydrogenase; AcCoA, acetyl coenzyme A; EMP, Embden–Meyerhof pathway; TCA, tricarboxylic acid cycle. The letter S indicates nonenzymatic spontaneous condensation reactions

TABLE 2 Information for proteases and peptidases relevant to the degradation of proteins and peptides in *Laceyella sacchari* FBKL4.010

FBKL4.010 encoding protein ID	Locus tags	Description of specific proteases and peptidases	Similarity to described protein and peptide degrading enzymes
AUS09177.1 AUS09178.1	C1X05_10310 C1X05_10315	HslVU is an ATP-dependent protease consisting of HslU and HslV subunits. HslU markedly stimulates the proteolytic activity of HslV which can slowly degrade specific hydrophobic peptides and polypeptides (Kang et al., 2014; Yoo et al., 1996)	98.95% with the HslU subunit in *Laceyella sediminis* 99.44% with the HslV subunit HslV in *Laceyella sacchari*
AUS09873.1	C1X05_14280	M32 carboxypeptidases are thermostable metalloproteases consisting of two members isolated from the thermophilic bacteria *Thermus aquaticus* (*Taq*CP) and *Pyrococcus furiosus* (*Pfu*CP). These enzymes hydrolyze peptides and sequentially release amino acids from the C-terminus, with a broad specificity toward a wide range of C-terminal substrates (Lee et al., 1996; Niemirowicz, Parussini, Fernán, & Cazzulo, 2007)	96.79% with the M32 Carboxypeptidase in *Laceyella sediminis*
AUS07652.1 AUS08399.1	C1X05_01430 C1X05_05860	Methionine aminopeptidases (MetAPs) are organized into two classes (types I and II), with type I MetAPs present in *Escherichia coli*. MetAP can remove N-terminal methionine residues from polypeptide chains and catalyze the release of several hydrophobic amino acids in addition to methionine (Mitra, Sheppard, Wang, Bennett, & Holz, 2009; Walker & Bradshaw, 2013)	99.60% with the type I methionyl aminopeptidase in *Laceyella sediminis* 98.80% with the type I methionyl aminopeptidase in *Laceyella sediminis*
AUS09934.1	C1X05_14610	Thermitase is a thermostable endoprotease with the ability to convert various food-relevant substrates into low-molecular-weight peptides. This enzyme can be produced by strains in *Laceyella* sp. and *Thermoactinomyces* sp. (Jørgensen, Madsen, Vrang, Hansen, & Johnsen, 2013)	99.74% with a thermitase member of the peptidase S8 family in *Laceyella sediminis*

encoding enzymes that are relevant to tetramethylpyrazine synthesis indicated the possible presence of the complete tetramethylpyrazine metabolic pathways in *L. sacchari* FBKL4.010.

Finally, we confirmed that all tetramethylpyrazine-relevant protein sequences encoded in the genome of FBKL4.010 showed the highest protein identity (≥95%) with other members of the family *Thermoactinomycetaceae* such as *L. sediminis* and *T. vulgaris* according to the annotation pipeline using GeneMarkS + v4.3 (Table A1).

In this study, metabolic pathways relevant to tetramethylpyrazine were also reconstructed based on the KEGG PATHWAY and Mapper databases. In the process of reconstructing the metabolic pathways for acetoin, we inferred that a complete Embden–Meyerhof pathway (EMP) exists in the genome of FBKL4.010. Pyruvate is an important precursor that is generated by EMP pathway. Thus, the metabolic pathway from pyruvate to acetoin was reconstructed (Figure A1). Although genes encoding ALS in the genome of FBKL4.010 genome were annotated in this pathway, those genes encoding α-acetolactate decarboxylase (ALDC) and 2,3-butanediol dehydrogenase (BDH) could not be confirmed. These results indicate that FBKL4.010 could possibly generate acetolactate. However, how the strain produces acetoin needs to be further tested.

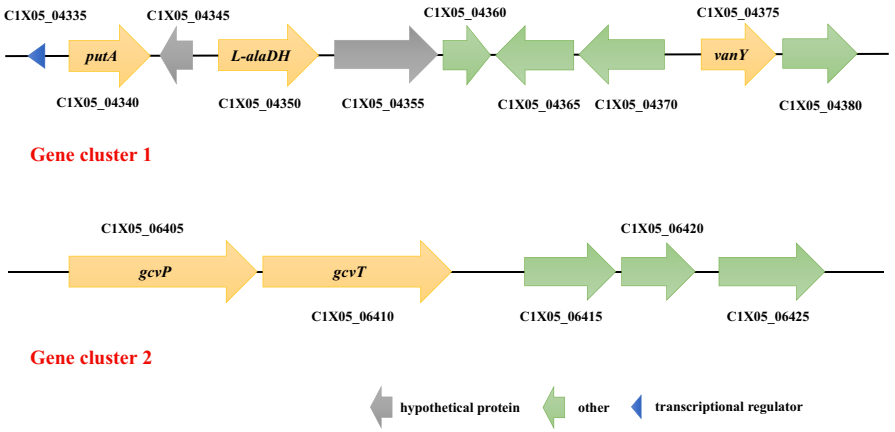

FIGURE 4 Two gene clusters related to key enzymes in the tetramethylpyrazine metabolic pathways. Lines among arrows represent intergenic spacers in clusters. The length of arrows and lines roughly reflects the gene and intergenic spacer length but not in actual proportions. Abbreviations: *putA*, proline dehydrogenase; *L-alaDH*, alanine dehydrogenase NAD-binding and catalytic domains; *vanY*, D-alanyl-D-alanine carboxypeptidase; *gcvP*, glycine dehydrogenase; *gcvT*, aminomethyl-transferring glycine dehydrogenase

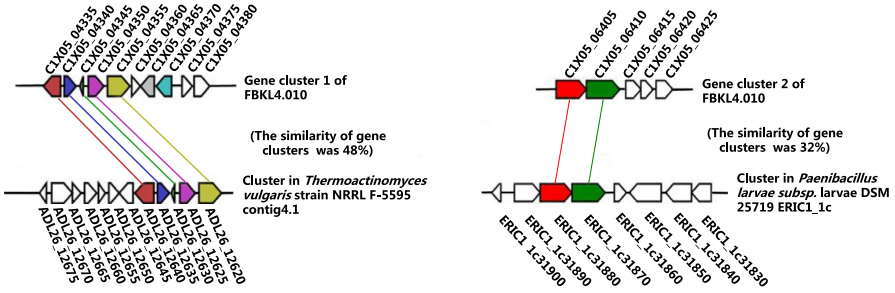

FIGURE 5 A synteny analysis to compare the alignment of the tetramethylpyrazine-related genes in two gene clusters between the strain FBKL4.010 and other closely related strains by antiSMASH 4.0. Each arrow represents a single gene. Homologous genes share a single color across strains. Vertical lines with corresponding colors connect homologous and conserved genes between the two clusters. Labels next to the arrows indicate gene position in the whole genome

Additionally, metabolic pathways of NH_3 production were reconstructed according to the information obtained for FBKL4.010. Two major metabolic pathways of NH_3 may exist in strain FBKL4.010. One pathway may start from L-glutamate and L-glutamine and finally generate NH_3 under the catalysis of glutamate dehydrogenase and glutamate synthase (Glevarec et al., 2004) (Figure A2). Another pathway may start at serine and glycine and generate NH_3 under the catalysis of serine hydroxymethyltransferase, glycine dehydrogenase, and aminomethyl-transferring glycine dehydrogenase (Alhasawi, Castonguay, Appanna, Auger, & Appanna, 2015; Zhang, Wu, & Chen, 2018) (Figure A3). From the above finding of several encoding genes relevant to protein, peptide degradation, and the production of free amino acids, it can be inferred that complete metabolic pathways of NH_3 production are present in strain FBKL4.010. Strain FBKL4.010 may produce NH_3 as an important precursor of tetramethylpyrazine via this metabolic pathway during the simulated fermentation of Moutai-flavor Daqu.

Two gene clusters differing from known tetramethylpyrazine synthesis clusters, encoding ADH and aminotransferase involved in tetramethylpyrazine metabolism, were also predicted in our study. Cluster 1 (from C1X05_04335 to C1X05_04380 loci) consists of ten genes (the *PutA* and *L-AlaDH* genes encoding ADH and the *vanY* gene encoding D-alanyl-D-alanine carboxypeptidase). Cluster 2 (from C1X05_06405 to C1X05_06425 loci) contains two aminotransferase-encoding genes (the *gcvP* and *gcvT* genes) (Figure 4). With the identification of two new gene clusters, we found that the distribution of some genes related to tetramethylpyrazine biosynthesis was regular. These results provide a guide for identifying other unconfirmed genes in tetramethylpyrazine metabolism from the whole-genome sequence of strain FBKL4.010. Furthermore, the order of tetramethylpyrazine-relevant genes in clusters 1 and 2 showed some similarity with other closely related strains, such as *T. vulgaris* NRRL F-5595 and *Paenibacillus larvae* DSM 25719 (Figure 5), and the similarity of the gene clusters was 48% and 32%, respectively. However, we did not find a similar alignment with genes in *B. subtilis* or other species that produce tetramethylpyrazine. This result may indicate that tetramethylpyrazine-relevant genes of the strain FBKL4.010 have been notably reordered.

4 | CONCLUSION

In conclusion, this is the first report on the complete genome of the tetramethylpyrazine-producing species *L. sacchari*. The whole-genome sequence analysis of *L. sacchari* FBKL4.010 showed that a set of protein-coding genes related to tetramethylpyrazine synthesis is

present in the genome of this strain. The results of this study provide an opportunity to further understand the tetramethylpyrazine synthesis pathway of *L. sacchari* in the process of liquor-making.

ACKNOWLEDGMENTS

This work was funded by grants of Industrial Research Project in Guizhou Province GZ [2014]3012 and the Special Project in Guizhou Province GZ [2015]6012. The authors are also really grateful to the Personal Biotechnology Co., Ltd, Shanghai, China, for sequencing the whole genome of *L. sacchari* FBKL4.010.

CONFLICT OF INTERESTS

None declared.

AUTHOR CONTRIBUTIONS

Shuyi Qiu, Dounan Li, Chunxiao Wang, and Wei Huang conceived the study. Shuyi Qiu, Dounan Li, Chunxiao Wang, and Wei Huang were involved in formal analysis. Shuyi Qiu acquired the funding. Shuyi Qiu and Dounan Li administrated the project. Shuyi Qiu provided resources. Shuyi Qiu and Dounan Li wrote the original draft of the manuscript. Shuyi Qiu, Dounan Li, Chunxiao Wang, and Wei Huang wrote, reviewed, and edited the manuscript. All authors read and approved the manuscript for publication.

ETHICS STATEMENT

None required.

DATA AVAILABILITY STATEMENT

Genome information for the chromosome of *Laceyella sacchari* FBKL4.010 is openly available from GenBank databases under the project number CP025943, and the link is https://www.ncbi.nlm.nih.gov/nuccore/CP025943. All data generated or analyzed during this study are included in this published article, and some supporting data are available in Appendix and in the Dryad repository at https://doi.org/10.5061/dryad.82gn3m6.

ORCID


Dounan Li https://orcid.org/0000-0002-9131-9783

REFERENCES

Alhasawi, A., Castonguay, Z., Appanna, N. D., Auger, C., & Appanna, V. D. (2015). Glycine metabolism and anti-oxidative defence mechanisms in *Pseudomonas fluorescens*. *Microbiological Research*, *171*, 26–31. https://doi.org/10.1016/j.micres.2014.12.001

Arndt, J. W., Hao, B., Ramakrishnan, V., Cheng, T., Chan, S. I., & Chan, M. K. (2002). Crystal structure of a novel carboxypeptidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Structure*, *10*, 215–224. https://doi.org/10.1016/S0969-2126(02)00698-6

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). Spades: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*, 455–477. https://doi.org/10.1089/cmb.2012.0021

Chen, J. J., Lin, L. B., Zhang, L. L., Zhang, J., Tang, S. K., & Li, W. J. (2012). *Laceyella sediminis sp*. nov. a thermophilic bacterium isolated from a hot spring. *International Journal of Systematic and Evolutionary Microbiology*, *62*, 38–42. https://doi.org/10.1099/ijs.0.028282-0

Delcher, A. L., Harmon, D., Kasif, S., White, O., & Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Research*, *27*, 4636–4641. https://doi.org/10.1093/nar/27.23.4636

Glevarec, G., Bouton, S., Jaspard, E., Riou, M. T., Cliquet, J. B., … Limami, A. M. (2004). Respective roles of the glutamine synthetase/glutamate synthase cycle and glutamate dehydrogenase in ammonium and amino acid metabolism during germination and post-germinative growth in the model legume medicago truncatula. *Planta*, *219*, 286–297. https://doi.org/10.2307/23388559

Hanphakphoom, S., Maneewong, N., Sukkhum, S., Tokuyama, S., & Kitpreechavanich, V. (2014). Characterization of poly (L-lactide)-degrading enzyme produced by thermophilic filamentous bacteria *Laceyella sacchari* LP175. *Journal of General and Applied Microbiology*, *60*, 13–22. https://doi.org/10.2323/jgam.60.13

Hershey, H. P., Schwartz, L. J., Gale, J. P., & Abell, L. M. (1999). Cloning and functional expression of the small subunit of acetolactate synthase from nicotiana plumbaginifolia. *Plant Molecular Biology*, *40*, 795–806. https://doi.org/10.1023/A:1006273224977

Huo, J., Huang, M., Sun, B., Zheng, F., Sun, J., & Li, H. H. (2017). Research progress in functional factors in Baijiu. *Liquor-Making Science & Technology*, *9*, 17–23. https://doi.org/10.13746/j.njkj.2017174

Jørgensen, C. M., Madsen, S. M., Vrang, A., Hansen, O. C., & Johnsen, M. G. (2013). Recombinant expression of *Laceyella sacchari*, thermitase in *Lactococcus lactis*. *Protein Expression and Purification*, *92*, 148–155. https://doi.org/10.1016/j.pep.2013.09.009

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*, D457–D462. https://doi.org/10.1093/nar/gkv1070

Kang, M. S., Lim, B. K., Seong, I. S., Seol, J. H., Tanahashi, N., … Chung, C. H. (2014). The ATP-dependent CodWX (HslVU) protease in *Bacillus subtilis* is an N-terminal serine protease. *European Molecular Biology Organization*, *20*, 734–742. https://doi.org/10.1093/emboj/20.4.734

Kaur, N., Arora, A., Kumar, N., & Mayilraj, S. (2014). Genome sequencing and annotation of *Laceyella sacchari* strain GS 1–1, isolated from Hot spring, Chumathang, Leh, india. *Genomics Data*, *2*, 18–19. https://doi.org/10.1016/j.gdata.2013.10.007

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*, 722–736. https://doi.org/10.1101/gr.215087.116

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., … Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*, R12. https://doi.org/10.1186/gb-2004-5-2-r12

Lee, S. H., Taguchi, H., Yoshimura, E., Minagawa, E., Kaminogawa, S., … Matsuzawa, H. (1996). The active site of carboxypeptidase Taq possesses the active-site motif His–Glu–X–X-His of zinc-dependent endopeptidases and aminopeptidases. *Protein Engineering*, *9*, 467–469. https://doi.org/10.1093/protein/9.6.467

Li, D. N., & Qiu, S. Y. (2017). Overview of microbial community structure and function in Moutai-flavor Daqu. *China Brewing*, *36*, 5–11. https://doi.org/10.11882/j.issn.0254-5071.2017.01.002

Li, H., Lian, B., Ding, Y., Nie, C., & Zhang, Q. (2014). Bacterial diversity in the central black component of maotai daqu and its flavor analysis. *Annals of Microbiology*, *64*, 1659–1669. https://doi.org/10.1007/s13213-014-0809-z

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., … Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience*, *1*, 1–6. https://doi.org/10.1186/2047-217X-1-18

Matsuo, Y., Katsuta, A., Matsuda, S., Shizuri, Y., Yokota, A., & Kasai, H. (2006). *Mechercharimyces mesophilus* gen. nov., sp. nov. and *Mechercharimyces asporophorigenens* sp. nov., antitumour substance-producing marine bacteria, and description of *Thermoactinomycetaceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, *56*, 2837–2842. https://doi.org/10.1099/ijs.0.64223-0

Mitra, S., Sheppard, G., Wang, J., Bennett, B., & Holz, R. C. (2009). Analyzing the binding of co(ii)-specific inhibitors to the methionyl aminopeptidases from *Escherichia coli* and *Pyrococcus furiosus*. *Journal of Biological Inorganic Chemistry*, *14*, 573–585. https://doi.org/10.1007/s00775-009-0471-2

Mount, D. W. (2007). Using the basic local alignment search tool (BLAST). *Cold Spring Harbor Protocols*, *14*, pdb.top17. https://doi.org/10.1101/pdb.top17

Niemirowicz, G., Parussini, F., Fernán, A., & Cazzulo, J. J. (2007). Two metallo carboxypeptidases from the protozoan *Trypanosoma cruzi* belong to the M32 family, found so far only in prokaryotes. *Biochemical Journal*, *401*, 399–410. https://doi.org/10.1042/BJ20060973

Pruitt, K. D., Tatusova, T., Klimke, W., & Maglott, D. R. (2009). Ncbi reference sequences: Current status, policy and new initiatives. *Nucleic Acids Research*, *37*, D32–D36. https://doi.org/10.1093/nar/gkn721

Shukla, R. J., & Singh, S. P. (2015). Characteristics and thermodynamics of α-amylase from thermophilic actinobacterium, *Laceyella sacchari*, TSI-2. *Process Biochemistry*, *50*, 2128–2136. https://doi.org/10.1016/j.procbio.2015.10.013

Stothard, P., & Wishart, S. D. (2005). Circular genome visualization and exploration using CGView. *Bioinformatics*, *21*, 537–539. https://doi.org/10.1093/bioinformatics/bti054

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., … Ostell, J. (2016). Ncbi prokaryotic genome annotation pipeline. *Nucleic Acids Research*, *44*, 6614–6624. https://doi.org/10.1093/nar/gkw569

Walker, K. W., & Bradshaw, R. A. (2013). Methionyl aminopeptidase type 1. In A. J. Barrett, N.D. Rawlings, & J. F. Woessner (Eds), *Handbook of proteolytic enzymes* (pp. 1495–1500). London: Academic Press. https://doi.org/10.1016/B978-0-12-382219-2.00336-7

Wang, C. L., Shi, D. J., & Gong, G. L. (2008). Microorganisms in Daqu: A starter culture of Chinese Maotai-flavor liquor. *World Journal of Microbiology & Biotechnology*, *24*, 2183–2190. https://doi.org/10.1007/s11274-008-9728-0

Wang, X. D., Ban, S. D., Zhou, H. X., Hu, B. D., & Qiu, S. Y. (2016). Comparative analysis of bacterial populations of three Maotai-Flavored Daqus in Zunyi, Guizhou. *Food Science*, *37*, 110–116. https://doi.org/10.7506/spkx1002-6630-201607021

Wang, X. D., Wang, J., Zhu, G. J., Lei, A. L., Hu, P. G., & Qiu, S. Y. (2017). Isolation and identification of functional bacteria with high yield of tetramethyl pyrazine from Moutai-flavor high-temperature Daqu. *China Brewing*, *36*, 55–60. https://doi.org/10.11882/j.issn.0254-5071.2017.01.011

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., … Medema, M. H. (2015). antiSMASH 3.0—A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, *43*, W237–W243. https://doi.org/10.1093/nar/gkv437

Whitman, W. B., Woyke, T., Klenk, H.-P., Zhou, Y., Lilburn, T. G., Beck, B. J., … Kyrpides, N. C. (2015). Genomic encyclopedia of bacterial and archaeal type strains, phase III: The genomes of soil and plant-associated and newly described type strains. *Standards in Genomic Sciences*, *10*, 26–31. https://doi.org/10.1186/s40793-015-0017-x

Wu, J. F. (2007). Study on functional composition in Chinese liquor–tetramethylpyrazine. *Liquor-making Science & Technology*, *1*, 117–120. https://doi.org/10.13746/j.njkj.2007.01.034

Wu, J. F., & Xu, Y. (2014). Formation mechanism of tetramethylpyrazine produced with *B. subtilis* S12 under the fermentation condition simulated bacterial Qu preparation used for Chinese liquor brewing. *Food Science and Biotechnology*, *33*, 8–15.

Xian, W. D., Ming, H., & Li, W. J. (2015). The family *Thermoactinomycetaceae* – A review. *Acta Microbiologica Sinica*, *55*, 1–11. https://doi.org/10.13343/j.cnki.wsxb.20140237

Xiao, Z., Hou, X., Xin, L., Xi, L., & Zhao, J. (2014). Accelerated green process of tetramethylpyrazine production from glucose and diammonium phosphate. *Biotechnology for Biofuels*, *7*, 106–114. https://doi.org/10.1186/1754-6834-7-106

Xiao, Z., Liu, P. H., Qin, J. Y., & Xu, P. (2007). Statistical optimization of medium components for enhanced acetoin production from molasses and soybean meal hydrolysate. *Applied Microbiology and Biotechnology*, *74*, 61–68. https://doi.org/10.1007/s00253-006-0646-5

Xu, Y., Wu, Q., Fan, W. L., & Zhu, B. F. (2011). The Discovery & verification of the production pathway of tetramethylpyrazine (TTMP) in Chinese liquor. *Liquor-Making Science and Technology*, *7*, 37–40. https://doi.org/10.13746/j.njkj.2011.07.001

Yang, F., Lin, L., Wang, H., Wang, L., Yang, D., … Ji, K. L. (2011). Comparative analysis of the metabolites of three *Bacillus* strains in Maotai daqu starter. *Liquor-Making Science & Technology*, *8*, 42–46. https://doi.org/10.13746/j.njkj.2011.08.030

Yao, S., Xu, Y. Q., Xin, C. H., Xu, L., Liu, Y., … Chen, C. (2014). Genome sequence of *Thermoactinomyces daqus* H-18, a novel thermophilic species isolated from high-temperature daqu. *Genome Announcements*, *3*, e01394-01414. https://doi.org/10.1128/genomeA.01394-14

Yoo, S. J., Seol, J. H., Shin, D. H., Rohrwild, M., Kang, M. S., … Chung, C. H. (1996). Purification and characterization of the heat shock proteins HslV and HslU that form a new ATP-dependent protease in *Escherichia coli*. *The Journal of Biological Chemistry*, *271*, 14035–14040. https://doi.org/10.1074/jbc.271.24.14035

Yoon, J. H., Kim, I. G., Shin, Y. K., & Park, Y. H. (2005). Proposal of the genus *Thermoactinomyces sensu* stricto and three new genera, *Laceyella*, *Thermoflavimicrobium* and *Seinonella*, on the basis of phenotypic, phylogenetic and chemotaxonomic analyses. *International Journal of Systematic and Evolutionary Microbiology*, *55*, 395–400. https://doi.org/10.1099/ijs.0.63203-0

Zhang, C. L., Ao, Z. H., Chui, W. Q., Shen, C. H., Tao, W. Y., & Zhang, S. Y. (2012). Characterization of the aroma-active compounds in Daqu: A tradition Chinese liquor starter. *European Food Research and Technology*, *234*, 69–76. https://doi.org/10.1007/s00217-011-1616-4

Zhang, M., Wu, W., & Chen, Z. (2018). Structure and function of cytoplasmic serine hydroxymethyltransferase from *pichia pastoris*. *Biochemical and Biophysical Research*, *496*, 753–757. https://doi.org/10.1016/j.bbrc.2018.01.084

Zhu, B. F., Xu, Y., & Fan, W. L. (2009). Study of tetramethylpyrazine formation in fermentation system from glucose by *Bacillus subtilis*, XZ1124. *New Biotechnology*, *25*, S237. https://doi.org/10.1016/j.nbt.2009.06.224

Zhu, B. F., Xu, Y., & Fan, W. L. (2010). High-yield fermentative preparation of tetramethylpyrazine by *Bacillus* sp. using an endogenous precursor approach. *Journal of Industrial Microbiology & Biotechnology*, *37*, 179–186. https://doi.org/10.1007/s10295-009-0661-5

How to cite this article: Li D, Huang W, Wang C, Qiu S. The complete genome sequence of the thermophilic bacterium *Laceyella sacchari* FBKL4.010 reveals the basis for tetramethylpyrazine biosynthesis in Moutai-flavor Daqu. *MicrobiologyOpen*. 2019;8:e922. https://doi.org/10.1002/mbo3.922

APPENDIX 1

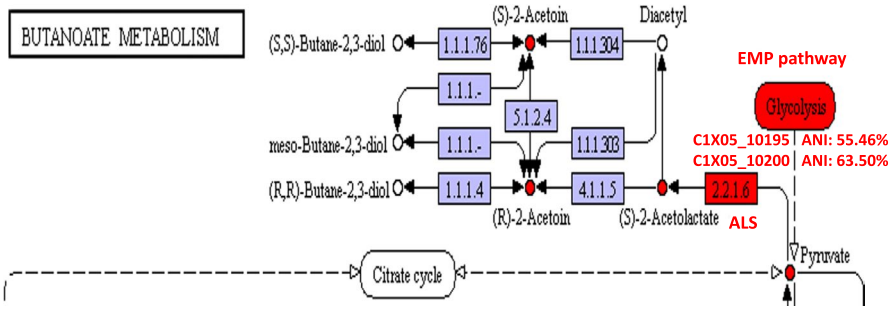

FIGURE A1 The reconstructed acetoin metabolic pathway in the strain FBKL4.010 based on the KEGG PATHWAY database. EMP represents Embden–Meyerhof pathway. Red marked genes represent genes related to tetramethylpyrazine synthesis on FBKL4.010 genome. The locus tags of marked genes and the ANI value with the most similar genes were shown next to the marked genes

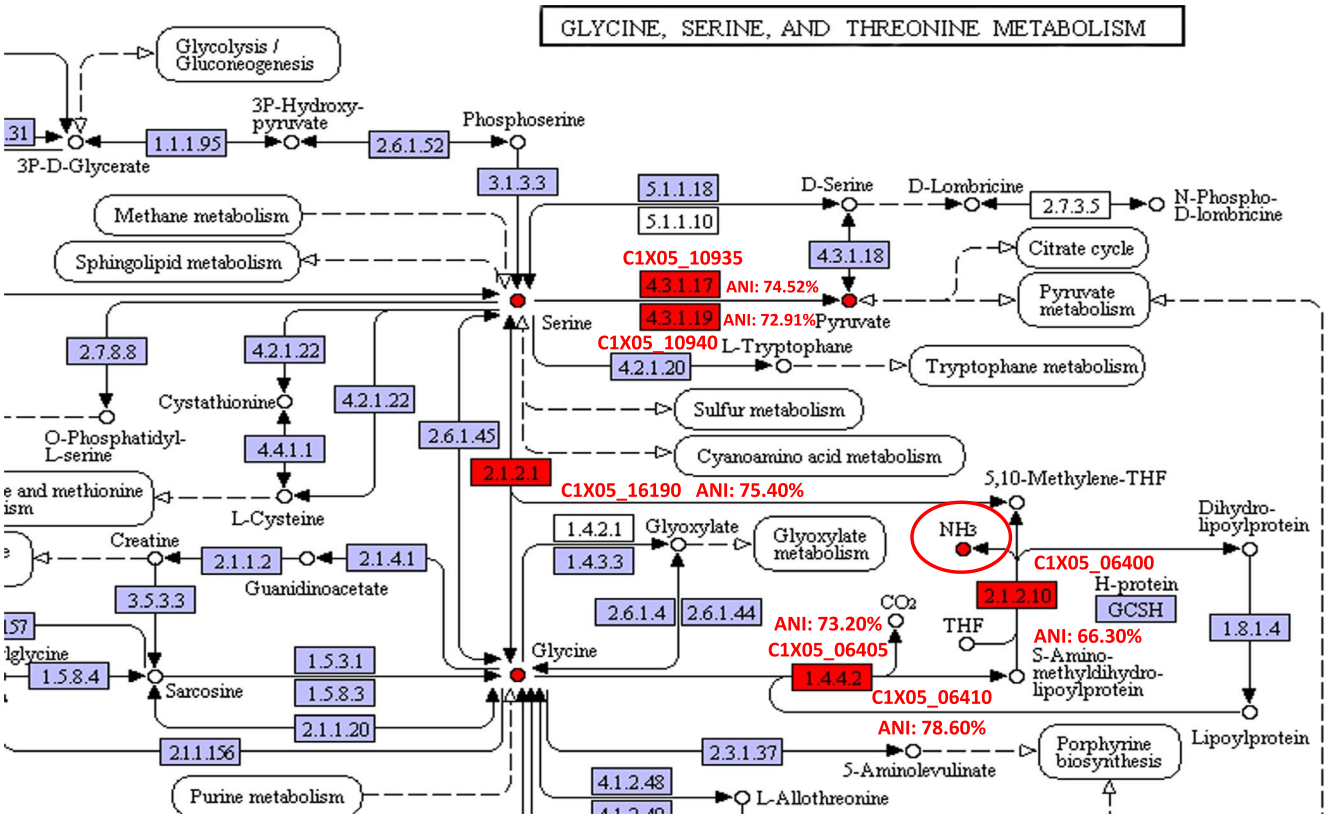

FIGURE A2 The reconstructed result of the ammonium metabolic pathway in the strain FBKL4.010 based on the KEGG PATHWAY database. Red marked genes represent genes related to tetramethylpyrazine synthesis on FBKL4.010 genome. The locus tags of marked genes and the ANI value with the most similar genes were shown next to the marked genes

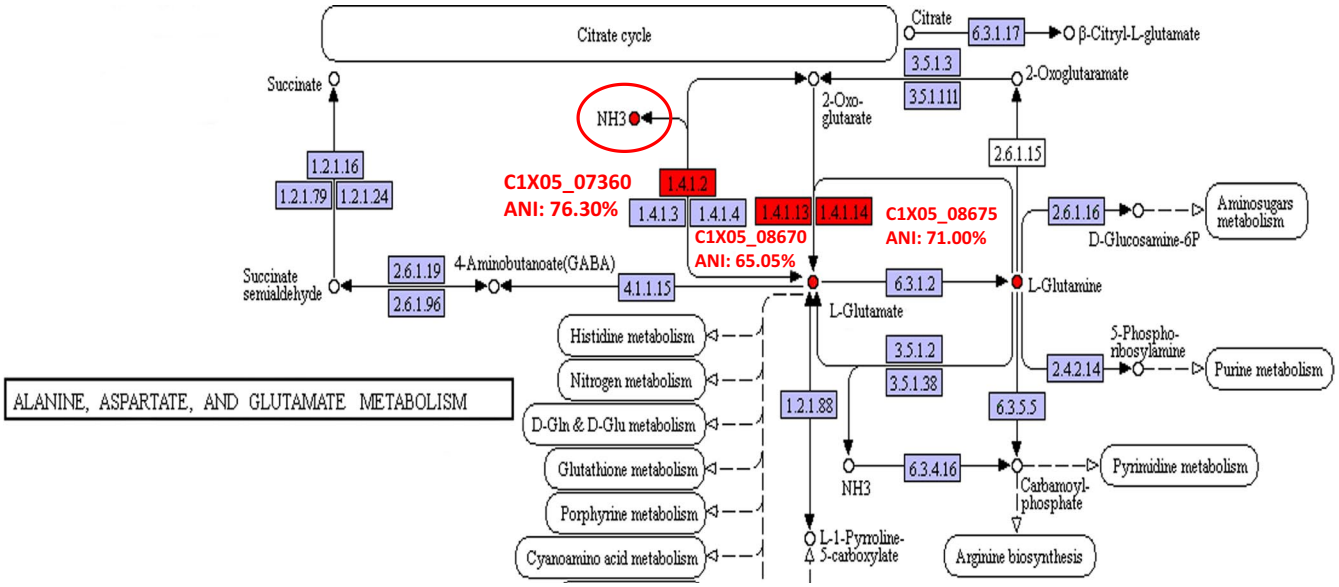

FIGURE A3 The reconstructed result of the ammonium metabolic pathway in the strain FBKL4.010 based on the KEGG PATHWAY database. Red marked genes represent genes related to tetramethylpyrazine synthesis on FBKL4.010 genome. The locus tags of marked genes and the ANI value with the most similar genes were shown next to the marked genes

TABLE A1 Protein sequences relevant to tetramethylpyrazine metabolism in *Laceyella sacchari* FBKL4.010

FBKL4.010 encoding protein ID	Locus tags	Most similar protein sequence ID	Similar family/genus/species	Identity %	Protein name
AUS10426.1	C1X05_08485	WP_106342963.1	*Laceyella sediminis*	99	Alpha-amylase (EC:3.2.1.1)
AUS09165.1	C1X05_10200	WP_106343376.1	*Laceyella sediminis*	98	Biosynthetic-type acetolactate synthase large subunit (EC:2.2.1.6)
AUS09164.1	C1X05_10195	WP_022739119.1	*Thermoactinomycetaceae*	95	Acetolactate synthase small subunit (EC:2.2.1.6)
AUS09177.1	C1X05_10310	WP_106342640.1	*Laceyella sediminis*	99	HslU-HslV ATP-dependent protease ATPase subunit
AUS09178.1	C1X05_10315	WP_022736267.1	*Laceyella sacchari*	99	HslU-HslV ATP-dependent protease peptidase proteolytic subunit
AUS09873.1	C1X05_14280	WP_106342307.1	*Laceyella sediminis*	97	Carboxypeptidase M32 (EC:3.4.17.-)
AUS07652.1	C1X05_01430	WP_106342798.1	*Laceyella sediminis*	99	Type I methionyl aminopeptidase (EC: 3.4.11.18)
AUS08399.1	C1X05_05860	WP_106341940.1	*Laceyella sediminis*	99	Type I methionyl aminopeptidase (EC: 3.4.11.18)
AUS09934.1	C1X05_14610	WP_106342352.1	*Laceyella sediminis*	99	Thermitase-like peptidase S8 (EC: 3.4.21.-)
AUS10364.1	C1X05_04375	WP_106343256.1	*Laceyella sediminis*	98	D-alanyl-D-alanine carboxypeptidase (EC: 3.4.17.14)
AUS08130.1	C1X05_04340	WP_102991771.1	*Laceyella* sp.	100	Proline dehydrogenase (EC: 1.5.5.2)
AUS08132.1	C1X05_04350	WP_106343258.1	*Laceyella sediminis*	99	Alanine dehydrogenase NAD-binding and catalytic domains (EC:1.4.1.1)
AUS08494.1	C1X05_06400	WP_054095413.1	*Thermoactinomyces vulgaris*	98	Glycine cleavage system aminomethyltransferase (EC: 2.1.2.10)
AUS08496.1	C1X05_06410	WP_054095412.1	*Thermoactinomyces vulgaris*	99	Aminomethyl-transferring glycine dehydrogenase (EC:1.4.4.2)
AUS08495.1	C1X05_06405	WP_106342890.1	*Laceyella sediminis*	99	Glycine dehydrogenase (EC:1.4.4.2)
AUS09283.1	C1X05_10935	WP_029071817.1	*Laceyella sacchari*	99	L-serine ammonia-lyase, iron-sulfur-dependent, subunit alpha (EC: 4.3.1.17)
AUS09284.1	C1X05_10940	WP_054096012.1	*Thermoactinomycetaceae*	99	L-serine ammonia-lyase, iron-sulfur-dependent, subunit beta (EC: 4.3.1.19)
AUS10223.1	C1X05_16190	WP_022737378.1	*Thermoactinomyces vulgaris*	99	Serine hydroxymethyl transferase (EC: 2.1.2.1)
AUS08578.1	C1X05_06835	WP_022736215.1	*Thermoactinomyces vulgaris*	99	Alanine dehydrogenase (EC:1.4.1.1)
AUS08555.1	C1X05_06715	WP_054095392.1	*Thermoactinomycetaceae*	99	Leucine dehydrogenase (EC:1.4.1.9)
AUS09758.1	C1X05_13650	WP_029071896.1	*Thermoactinomyces vulgaris*	99	Leucine dehydrogenase (EC:1.4.1.9)
AUS08677.1	C1X05_07360	WP_022738428.1	*Thermoactinomyces vulgaris*	99	Glutamate dehydrogenase (EC:1.4.1.3)
AUS08910.1	C1X05_08670	WP_106342055.1	*Laceyella sediminis*	99	Glutamate synthase large subunit (EC: 1.4.1.13)
AUS08911.1	C1X05_08675	WP_054096311.1	*Thermoactinomyces vulgaris*	99	Glutamate synthase (EC: 1.4.1.14)