Xue *et al. Annals of Microbiology* (2023) 73:16
https://doi.org/10.1186/s13213-023-01719-6

Annals of Microbiology

UNIVERSITÀ DEGLI STUDI DI MILANO

REVIEW ARTICLE

Open Access

Reveal the microbial communities and functional prediction during the fermentation of Fen-flavor *Baijiu* via metagenome combining amplicon sequencing

Teng-da Xue[1], Jin-hua Zhang[1,2], Tian-rui Wang[3], Bao-qing Bai[1,2], Zhi-xing Hou[1], Jian-feng Cheng[4], Tao Bo[5*] and San-hong Fan[1,2*]

Abstract

Purpose Microbial resources are abundant in fermented grains of the Chinese Fen-flavor *Baijiu*, which is closely related to the quality of *Baijiu*. The purpose of this study was to investigate the microbial community structure and function in *Daqu* and fermented grains.

Methods We systematically compared two technical approaches, amplicon sequencing, and metagenomic sequencing, to analyze the microbial communities during *Baijiu* fermentation.

Result The results showed that lactic acid bacteria (LAB) and yeasts were the main microorganisms in the fermentation process. Firmicutes (*Lactobacillus*, *Pediococcus*, and *Weissella*) were the dominant bacteria, and Ascomycota (*Issatchenkia* or *Pichia*) was the dominant fungus in fermented grains. Moreover, *Pichia kudriavzevii*, *Lichtheimia ramosa*, and *Companilactobacillus paralimentarius* were the dominant species at the initial stage of fermentation by metagenomic sequencing. *Latilactobacillus curvatus*, *Loigolactobacillus coryniformis* subsp. *coryniformis*, and *Lentilactobacillus parabuchneri* became dominant during the middle stage of fermentation. *Lentilactobacillus parabuchneri* and *Lactobacillus acetotolerans* were the dominant species in the final stage of fermentation. Spearman correlation analysis showed that LAB inhibited the growth of yeasts.

Conclusion Combining the two sequencing methods provided valuable insights into the dynamic succession of microorganisms during the fermentation of *Baijiu*. It had had a particular significance for mining microbial species resources in fermented grains.

Keywords Metagenomics, Amplicon sequencing, *Baijiu*, Microbial community, Spearman correlation

*Correspondence:
Tao Bo
botao@sxu.edu.cn
San-hong Fan
fsh729@sxu.edu.cn
Full list of author information is available at the end of the article

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Introduction

Chinese traditional *Baijiu* is one of the oldest distilled spirits in the world (Xu et al. 2017). *Baijiu* can be divided into 12 types based on aroma characteristics (Xu et al. 2022), and each type of *Baijiu* is produced by its unique processing technology and complex fermentation system. The Fen flavor (light flavor) is one of the basic flavor types of *Baijiu* (Hong et al. 2020; Xu et al. 2017). Fen-flavor *Baijiu* is distilled after spontaneous fermentation of fermented grains (*Jiupei*) under a specific environment and unique process, and ethyl acetate is the hallmark aroma component (Wu et al. 2021). Fermented grain is formed by mixing *Daqu* and sorghum in a certain proportion. *Daqu* is a unique multi-strain starter for Chinese *Baijiu*, produced by solid-state fermentation from barley and pea (Huang et al. 2020a, b; Zheng et al. 2012). It not only provides functional microorganisms for the fermentation of *Baijiu* but also provides a variety of enzymes for the saccharification and fermentation of fermented grains and participates in the metabolism of flavor substances (Wang et al. 2020; Hou et al. 2022; Y. Huang et al. 2017a, b). Therefore, *Daqu* is being exploited as a new microbial resource (Huang et al. 2017a, b). In fermentation, the microbial communities in fermented grains adapt to the environment, continuously undergo dynamic succession, decompose raw materials during growth and metabolism and generate a variety of compounds in the fermentation system (Kong et al. 2014; Pang et al. 2020). At the same time, raw materials regulate flavor formation by driving microorganisms in *Baijiu* fermentation (Liu et al. 2019a, b). Therefore, the composition and succession of microbial communities in the fermentation system of *Baijiu* are directly related to the quality of *Baijiu*. Studying microbial community structure in *Daqu* and fermented grains has been a critical topic in *Baijiu* research.

The emergence of next-generation sequencing (NGS) technology has been used to explore microbial structure in a variety of sample systems. It can predict microbial function through gene annotation (Wensel et al. 2022), which makes up for the limitations of traditional culture methods in the study of microbial community structure (He et al. 2017). Amplicon sequencing and metagenomic sequencing are the most widely used next-generation sequencing technologies (Chakravorty et al. 2007; Knight et al. 2018), which have been used to study microorganisms that are closely related to the quality of Chinese *Baijiu*, including *Daqu* (Zhang et al. 2022); fermentation pit muds of *Baijiu* (Xu et al. 2020); fermented grains (Wang et al. 2022); and fermentation environment (Ma et al. 2022). At present, some studies have used amplicons sequencing and metagenomic sequencing to dissect microorganisms in the fermentation of Fen-flavor *Baijiu* (Huang et al. 2020a, b; Kang et al. 2022a). The two sequencing methods have similar results in the analysis of microbial diversity (Zuo et al. 2022). However, amplicon sequencing can only obtain structures for bacteria or fungi, respectively (Davis et al., 2016). Metagenome sequencing can understand the structure of microorganisms (bacteria and fungi) in a sample and identify microbial species (species or subspecies) with higher accuracy and predict their functions (Knight et al. 2018; Zhao & Eun 2020). The combination of amplicon and metagenomic sequencing can better interpret microbial community structure and function.

In this study, amplicon sequencing and metagenome sequencing were combined to reveal the microbial diversity succession, sample differential microorganisms, symbiosis, and potential metabolic functions during the fermentation of Fen-flavor *Baijiu*. This study provided information for exploring and utilizing microbial resources in fermented grains of traditional Chinese *Baijiu*. Then, the correlation between microorganisms and metabolic components in *Baijiu* could be studied, which was of great significance for the improve quality of *Baijiu*.

Materials and methods

Sampling

Daqu and fermented grains of *Baijiu* were collected from Shanxi *Fenjiu* Distillery (Fenyang, Shanxi, China). All samples, including *Daqu* and fermented grains, were from the same production batch. Samples were collected from *Daqu*, early, middle, and late stages of fermentation, i.e., at 0, 7, 15, and 28 days, the pH values were 6.65, 5.21, 4.67, 3.80 and 3.78, respectively. Sample collection and processing methods were as follows: firstly, samples obtained at different time points were from the same earthenware jars. Secondly, positions were fixed for each sampling, and samples were quantitatively collected at 50 cm below the surface of the fermented grains and mixed well. Next, the samples were transported back to the laboratory and refrigerated at − 80 °C. The *Daqu* sample and the fermented grains samples were divided into three samples for parallel testing, and they were labelled as JQ, JP0, JP7, JP15, and JP28, respectively.

DNA extraction and inspection

The genomic DNA of the samples was extracted by the CTAB method, as previously described (Zhang et al. 2022). The quality of genomic DNA was assessed with 1% agarose gel. DNA concentration was measured using Qubit® dsDNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). OD260/OD280 value was between 1.8 ~ 2.0; DNA contents above

1 μg are used to construct the library. Amplicon and metagenomic sequencing were completed at Beijing Novogene, Beijing, China.

Amplicon sequencing

PCR amplification and sequencing

Bacterial 16S rDNA genes were amplified by PCR using primers 341F/806R with barcode (Liu et al. 2019a, b), and fungal ITS genes were amplified using primers ITS1F/ITS2 (Li et al. 2020). The diluted genomic DNA was used as a template. PCR was performed with Phusion® High-Fidelity PCR Master Mix with GC buffer and high-efficiency high-fidelity enzyme (New England Biolabs, Beijing, China) to ensure amplification efficiency and accuracy. The PCR products were detected by electrophoresis in 2% agarose gel, and the samples were mixed in equal amounts according to the concentration of PCR products. After thoroughly mixing, the samples were detected again by electrophoresis in 2% agarose gel, and the target bands were recovered by gel recovery kit (Qiagen, Hilden, Germany). NEBNext® Ultra™ II DNA Library Prep Kit was used to construct the library, and the constructed library was subjected to Qubit and Q-PCR quantification. The qualified libraries were sequenced using NovaSeq 6000.

Sequencing data processing and analysis

Each sample data was split from the sequencing data, and the reads of the sample were spliced to obtain raw tags and quality control to obtain high-quality clean tags. Chimera filtering was performed on the clean tags to obtain effective tags that could be used for subsequent analysis. The effective tags were denoised by the DADA2 module in QIIME2 software, and sequences with an abundance of less than five were filtered out to obtain the final amplicon sequence variants (ASVs) and feature sheets (Callahan et al., 2016). For the obtained ASVs (Callahan et al., 2017), on the one hand, species annotation was performed on the representative sequence of each ASV to obtain the corresponding species information and species-based abundance distribution, as well as species richness and evenness information within the sample. On the other hand, multi-sequence alignment of ASVs was performed, and nonmetric multidimensional scaling (NMDS) was used to explore the differences in community structure among different samples. *T*-test statistical analysis methods were used to test the significance of differences in species composition and community structure. Spearman correlation analysis was used to analyze the interaction between different microorganisms.

Metagenomic sequencing

Library construction and sequencing

A total amount of 1 μg DNA per sample was used as input material for the DNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Briefly, the DNA sample was fragmented by sonication to a size of 350 bp. Then, DNA fragments were end-polished, A-tailed, and ligated with the full-length adaptor for Illumina sequencing with further PCR amplification. At last, PCR products were purified (AMPure XP system), and libraries were analyzed for size distribution by Agilent 2100 Bioanalyzer and quantified using real-time PCR.

The index-coded samples were clustering on a cBot Cluster Generation System according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq PE150 platform, and paired-end reads were generated.

Sequencing data processing and analysis

We were preprocessing the Raw Data from the Illumina HiSeq sequencing platform using Readfq software to obtain a Clean Date. Metagenome assembly of clean data was performed by SOAPdenovo software. An abundance of information on each sample in the gene catalogue was obtained. Species annotation information of UniGene was obtained from the gene catalogue and microRNA database. Combined with the gene abundance table, species abundance tables at different taxonomic levels were obtained. We compared the gene sequences to obtain functional information with two available databases: Kyoto Encyclopedia of Genes and Genomes (KEGG) and carbohydrate-active enzymes (CAZy).

Results

Amplicon sequencing

Microbial community diversity

Figure 1 shows the relative abundance of microorganisms at the phylum and genus level for bacteria and fungi based on amplicon sequencing. Firmicutes was the dominant phylum in all samples at the phylum level of bacteria (Fig. 1A). In *Daqu*, the dominant phyla were mainly Firmicutes (69%), Proteobacteria (7%), Cyanobacteria (4%), and Bacteroidota (4%). At the beginning of fermentation (JP0), Firmicutes, Proteobacteria, and Cyanobacteria in fermented grains were provided by *Daqu*. Actinobacteria was probably provided by raw materials such as sorghum, as it only was found in sample JP0 and was detected in *Daqu* and fermented grains at the middle and late stages of fermentation. The relative abundance of Firmicutes

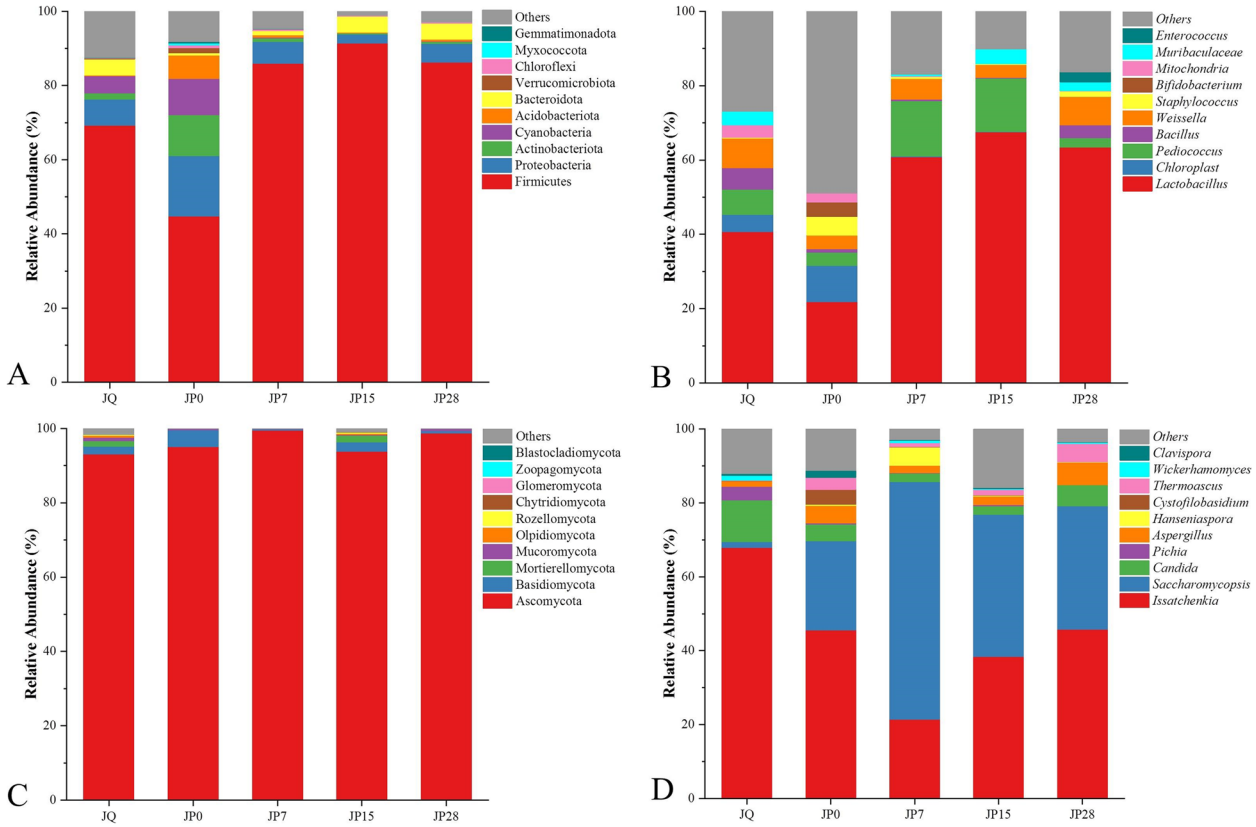

Fig. 1 Microbial community diversity based on amplicon sequencing (top 10 in relative abundance). Bacterial diversity at phylum (**A**) and genus level (**B**). Fungal diversity at phylum (**C**) and genus level (**D**)

increased to more than 80% as fermentation progressed. The relative abundance of Proteobacteria decreased in the early stage (0–7 days) and middle stage (7–15 days) of fermentation but increased in the late stage (15–28 days) of fermentation. The abundance of Bacteroidota continued to increase during fermentation.

At the genus level of bacteria (Fig. 1B), *Lactobacillus* was the dominant genus in all samples, followed by *Pediococcus* and *Weissella*. The relative abundance of *Lactobacillus* continued to increase as fermentation progressed, with the most incredible abundance at 15 days of fermentation and a slight decrease at the late stage of fermentation, probably due to the higher ethanol content in the late stages, which was not suitable for the growth of *Lactobacillus*. *Pediococcus* increased in relative abundance in the early stages of fermentation and remained stable during the middle of fermentation but decreased during the late stages. *Weissella* was present throughout the whole fermentation stage, and the relative abundance was relatively stable in the early and middle stages, with a slight increase in the late fermentation stage. The abundance of *Bacillus* in JQ and JP28 was more significant than 1%, and the relative abundance was very low in the early and middle stages of fermentation. *Staphylococcus* was detected at the beginning of fermentation (JP0) but not in *Daqu* (JQ), indicating that *Staphylococcus* was mainly derived from sorghum. As fermentation progressed, the relative abundance of *Staphylococcus* decreased and was detected only in JP28. *Enterococcus* was only detected in the 28-day fermentation sample (JP28). The above *Lactobacillus*, *Pediococcus*, *Weissella*, and *Enterococcus* all belonged to lactic acid bacteria (LAB), indicating that LAB was the dominant bacteria in both the *Daqu* and fermented grain samples.

At the phylum level of fungi (Fig. 1C), Ascomycota was the dominant fungus in both *Daqu* and the whole fermentation process, with relative abundance above 90%. At the genus level (Fig. 1D), *Issatchenkia* (68%), *Candida* (11%), and *Pichia* (4%) were the primary dominant fungi in *Daqu*, while *Pichia* had a relative abundance of less than 1% during fermentation. *Issatchenkia* and *Saccharomycopsis* are the dominant fungi during fermentation. The relative abundance of *Issatchenkia* decreased from 46 to 21% at the early stage of the fermentation

(0–7 days), after which it gradually increased. However, the relative abundance of *Saccharomycopsis* increased rapidly from 24% on day 0 to 64% on day 7 and then decreased continuously. In addition, *Candida* and *Aspergillus* were present in each sample, with a higher relative abundance in JP0 and JP28 than in JP7 and JP15. *Hanseniaspora* increased abundance during the early fermentation stage (0–7 days) and was not detected during the middle and late stages. In summary, yeasts (*Issatchenkia*, *Candida*, *Pichia*, *Saccharomycopsis*, *Hanseniaspora*, *Cystofilobasidium*, and *Wickerhamomyces*) were the predominant fungi in *Daqu* and fermented grains.

Previous studies have shown that the dynamic succession of microbial communities is related to many factors, among which biological factors (intraspecies genomic diversity and microbial structure) (Tan et al. 2022), fermentation parameters (starch, moisture, acidity, and reducing sugar) (Wang et al. 2021), and fermentation environment (container, factory, climate, ecology) (Kang et al. 2022b; Wang 2022) and other factors are considered to drive the succession of core and distinct microbial communities.

Alpha-diversity index analysis

We compared the α-diversity of microbiota between *Daqu* and fermented grains using the Chao1, pielou_e, Shannon, and Simpson index (Fig. 2) to evaluate changes in microbial diversity. The diversity of bacteria was higher than that of fungi throughout fermentation. The JP0 sample had a significantly higher than that of the other samples. The Chao1 and Shannon indices of bacteria decreased significantly first and then increased slightly, while the pielou_e and Simpson indices of bacteria did not change significantly. Shannon and Simpson of fungi first decreased significantly in the early stage of fermentation, increased significantly in the middle stage, and decreased in the late stage.

Beta-diversity analysis

Nonmetric multidimensional scaling (NMDS) (Kruskal 1964) was used to analyze the degree of difference between the samples, which was reflected by the distance between points. If the stress is less than 0.2, NMDS can accurately reflect the degree of difference between samples. The result is shown in Fig. 3. For both bacteria and fungi, there was a significant difference between one replicate sample of *Daqu* and three replicates of JP0, while the replicate samples of JP7, JP15, and JP28 were close. Moreover, the community structure of samples in the middle and late fermentation stages was similar, indicating that the microbial community structure in fermented grains gradually stabilized with the fermentation process.

Microbial difference analysis

T-test was used to analyze the microbial differences in different samples and the species with significant

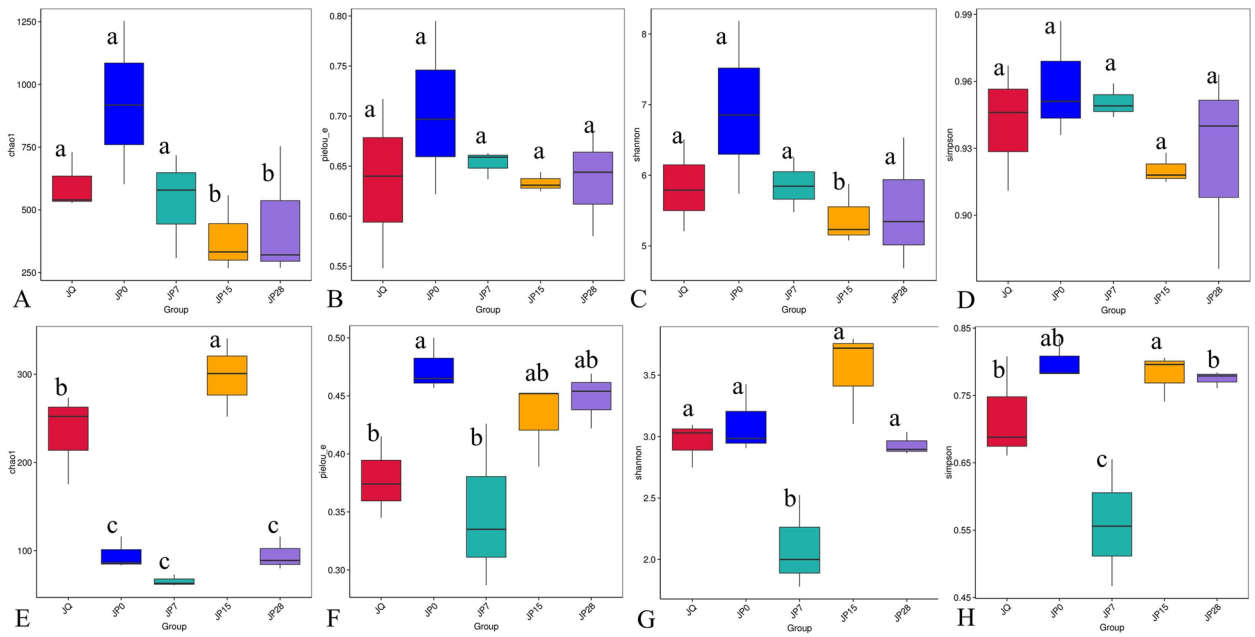

Fig. 2 Alpha-diversity index analysis of bacteria (**A**–**D**) and fungi (**E**–**H**). [a–c]Mean values in the same index with different lowercase letters differ significantly ($p < 0.05$)

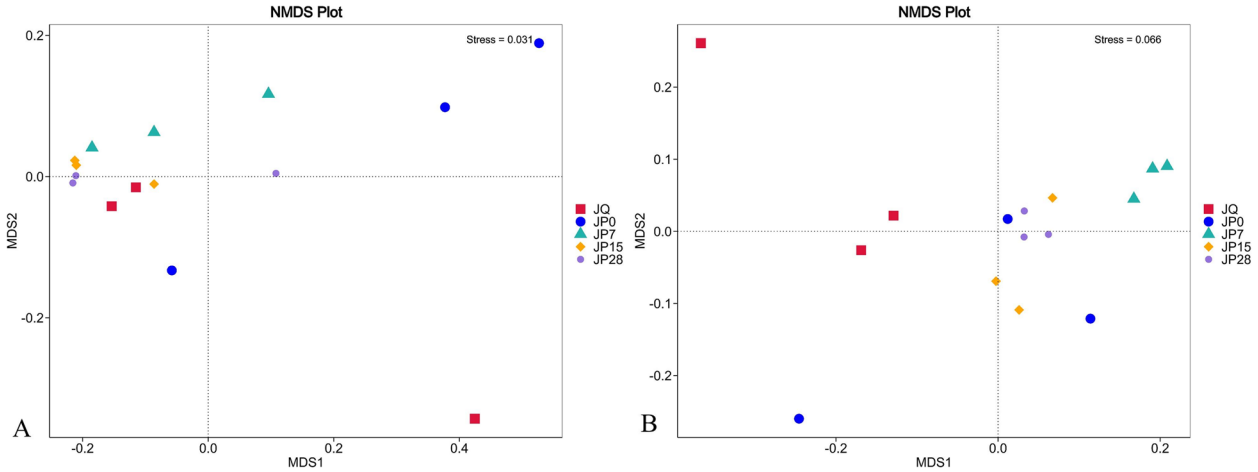

Fig. 3 NMDS analysis of bacteria (**A**) and fungi (**B**) community. A closer distance between points indicates higher similarity; $n = 3$

differences among the groups at each taxonomic level ($p < 0.05$). The analysis results between bacteria and fungi at the genus level are shown in Fig. 4. *Weissella*, *Cladosporium*, and *Fusarium* in *Daqu* were significantly more abundant than in JP0 samples. At the early stage of fermentation (JP0 vs. JP7), the abundance of *Pediococcus*, Muribaculaceae, and *Saccharomycopsis* increased significantly. At the middle stage of fermentation (JP7 vs. JP15), there was no significant change in bacteria; *Issatchenkia* increased significantly, while *Saccharomycopsis* decreased significantly. At the late fermentation stage (JP15 vs. JP28), the abundance of *Pediococcus*, *Acetobacter*, *Leuconostoc*, and other bacteria decreased significantly, while *Staphylococcus* increased significantly. However, *Aspergillus*, *Thermoascus*, *Rhizopus*, and other fungi increased significantly, and *Hanseniaspora* decreased significantly.

Correlation network of bacterial and fungal community

Respectively choose the relative abundance of bacteria and fungi in the level before the results of 40, adopt Spearman correlation analysis of testing, and select the relevant coefficient $|R| > 0.6$ and $p < 0.05$ microorganisms were plotted, and the results are shown in Fig. 5. In the correlation network diagram of bacteria (Fig. 5A), *Lactobacillus* was negatively correlated with most bacteria, indicating that the growth of *Lactobacillus* could inhibit the growth of other bacteria during the fermentation process. In addition, *Pediococcus* was positively correlated with *Weissella* and negatively correlated with *Bacillus*. In the correlation network diagram of fungi (Fig. 5B), *Issatchenkia* was found to be negatively correlated with *Saccharomycopsis* and *Hanseniaspora*, as well as *Candida* and *Hanseniaspora*.

Metagenomics sequencing

Microbial community structure

The microbial community structure in the samples based on metagenomic sequencing is shown in Fig. 6. At the phylum level (Fig. 6A), Firmicutes, Ascomycota, and Mucoromycota were the dominant phylum in *Daqu* and fermented grains. Among them, the relative abundance of Ascomycota in *Daqu* was 39%, which gradually decreased during the fermentation process, reaching the lowest (4%) on the 15th day of fermentation (JP15), and increased slightly in the late stage of fermentation. The relative abundance of Firmicutes increases during the early and middle stages of fermentation and decreases somewhat during the late stages.

At the genus level (Fig. 6B), *Lactobacillus* (17.4%), *Pichia* (33.2%), *Lichtheimia* (6.5%), and *Leuconostoc* (2.1%) were the predominant microorganisms in *Daqu*. At the beginning of fermentation (JP0), the relative abundance of *Lactobacillus*, *Pichia*, and *Lichtheimia* was 20.7%, 24.0%, and 13.1%, respectively, and the microbial structure was similar to that of *Daqu*. During the fermentation process, the relative abundance of *Lactobacillus* and *Pediococcus* reached a maximum on the 15th day of fermentation (JP15) and decreased slightly at the late fermentation stage. *Leuconostoc* was highest on the 7th day of fermentation and then reduced gradually. The relative abundance of *Pichia* and *Lichtheimia* decreased during fermentation's early and middle stages and increased slightly during the late stages. *Saccharomyces* showed an increasing trend during the whole fermentation process. Overall, the rapid growth of LAB (*Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Weissella*) inhibited the growth of yeast (*Pichia*, *Saccharomyces*, *Candida*) and molds (*Lichtheimia*, *Rhizopus*) in the early and middle stages of fermentation, while in the late stage of fermentation,

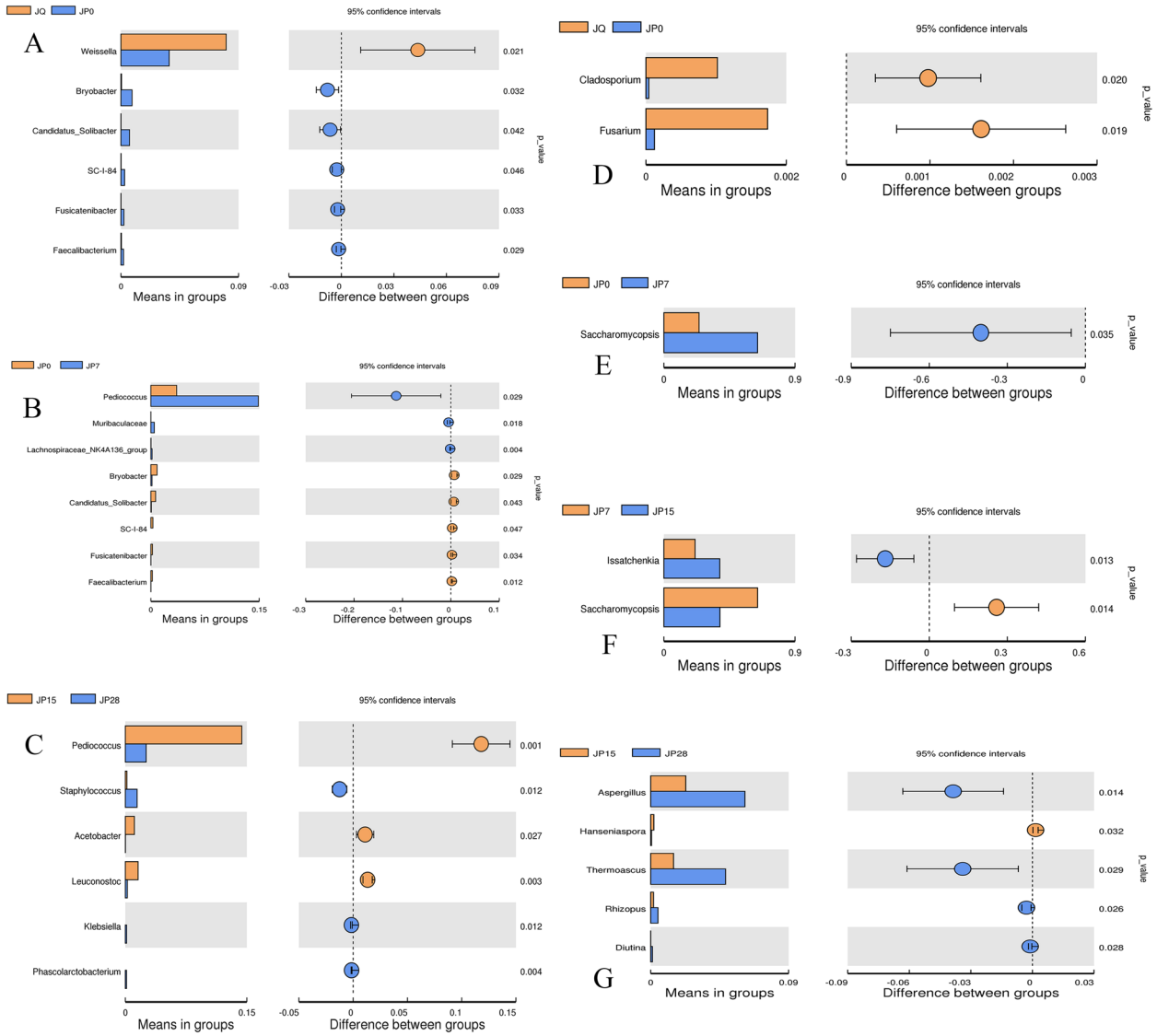

Fig. 4 *T*-test analysis of differences of bacteria (**A**–**C**) and fungi (**D**–**G**) at genus level in different fermentation stages. The bacteria in JP7 and JP15 samples were not significantly different

the fermentation tended to be complete with LAB was slightly reduced, and yeast increased somewhat.

Combining the clustering heat map of the top 35 microorganisms in terms of abundance at the genus level (Fig. 6C) revealed that the diversity of microbial populations gradually decreased as the fermentation proceeded. Still, the abundance of the dominant bacteria increased, with *Lactobacillus*, *Pediococcus*, and *Saccharomyces* mainly acting as fermenters in the late stages of fermentation.

In conclusion, microorganisms at the beginning of fermentation were mainly from *Daqu*, and yeasts were more abundant than LAB and were the dominant species in fermented grains. In the early and middle stages of fermentation, microbial succession in fermented grains was significant, the abundance of yeast decreased, and LAB continued to increase. At the later stage of fermentation, the microbial community structure tended to be stable, LAB became the main dominant species, and the relative abundance of yeast increased slightly.

Figure 7 shows the microbial community structure in each sample, with inner to outer circles showing species abundance at phylum, genus, and species levels, respectively. In JQ and JP0, *Pichia kudriavzevii* and *Lichtheimia ramosa* were the main dominant fungi, and *Companilactobacillus paralimentarius* was the main dominant

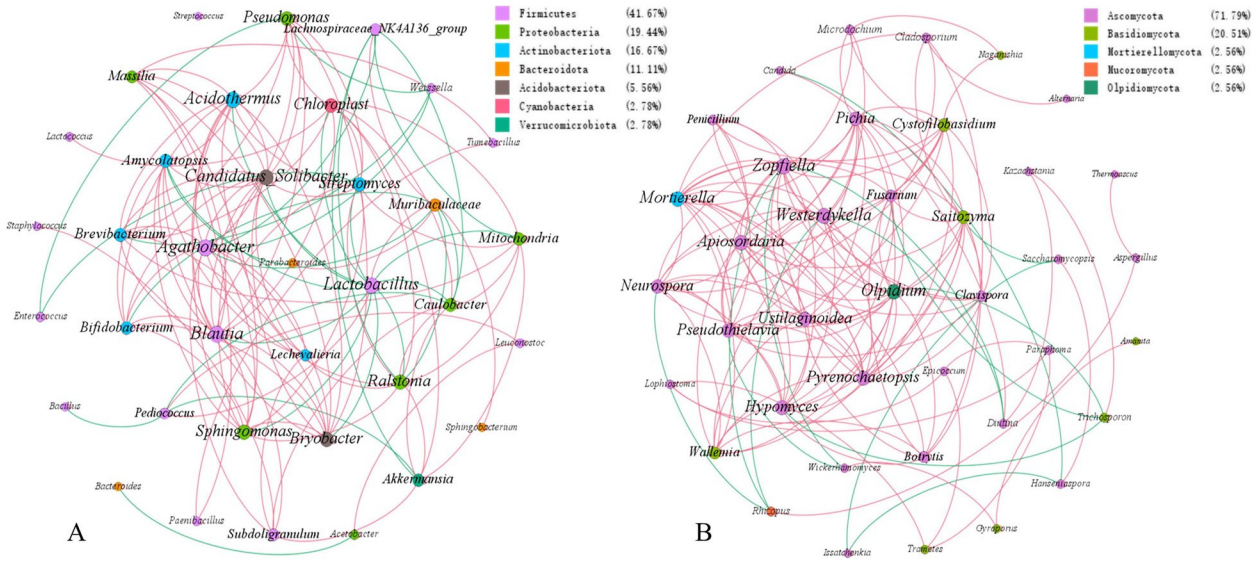

Fig. 5 Correlation networks of bacteria (**A**) and fungi (**B**) based on amplicon sequencing. Different nodes represent different genera, node size represents the degree of connectivity of the genus, the same color represents the level of the same phylum, the thickness of the line represents the size of correlation, the red line indicates positive correlation, and the blue line indicates negative correlation

bacteria. With the fermentation process, fungi's abundance rapidly decreased, while bacteria's abundance increased, and continuous succession occurred. At the genus level, a fermentation system with *Lactobacillus* as the main dominant bacteria was gradually formed, while at the species level, the species was still undergoing succession. In the late fermentation stage, the abundance of *Lactobacillus acetotolerans* increased from 0.2 to 9% and became the dominant species together with *Lentilactobacillus parabuchneri*. Huang et al. (2020a, b) found that *Lactobacillus acetotolerans*, *Lentilactobacillus buchneri* subsp. *buchneri*, and *Lentilactobacillus hilgardii* were the main dominant strains in fermented grains. Wang and Xu (2019) used a culture-dependent method to find that *Pichia anomala* and *Saccharomyces cerevisiae* were the dominant yeast species in Fen-flavored *Baijiu*. Different from the dominant species in this study, all strains were detected, probably due to different production batches, resulting in differences in species at the species level.

Correlation network of microbial community

Metagenomics sequencing provides information on all the microorganisms, based on the species abundance by the Spearman correlation analysis between total microbial relationship, select the relevant coefficient $|R|>0.6$, $p<0.05$ for drawing of microorganism; the result is shown in Fig. 8. The result is 34 nodes, 196 edges, 181 negative correlations (red lines), and 15 positive correlations (blue lines). *Lactobacillus* was negatively correlated with *Pichia*, *Bacillus*, and *Pseudomonas*, while *Candida* was positively correlated with *Lactococcus* and *Weissella*. In conclusion, during the fermentation of fermented grains, lactic acid and bacteriocin produced by LAB inhibited the growth and metabolism of some microorganisms (Porto et al. 2017), and antagonistic relationships existed among different yeasts, resulting in changes in microbial community structure.

Functional gene annotation by blasting to KEGG databases

KEGG and CAZy databases are using annotated metagenomic data to explore the functional characteristics and differences of microorganisms in *Daqu* and fermented grains at different fermentation stages. Figure 9 shows the results of functional gene annotation based on the KEGG database. Metabolism is the functional gene with the most significant number of annotations (Fig. 9A), which was the most abundant gene in *Daqu* and fermented grains, and its abundance reached 15.5% at day 15 of fermentation (Fig. 9B). According to the cluster heat map of available abundance under KEGG level 2 classification (Fig. 9C), the function of *Daqu* (JQ) is similar to that at the beginning of fermentation (JP0), and energy metabolism had the highest relative abundance. During fermentation, carbohydrate and amino acid metabolism increased and became the central functional gene. In addition, membrane transport was the primary functional gene in fermentation's middle and late stages.

In the metabolic pathway analysis at level 3 of KEGG (Fig. 9D), 396 categories were annotated. At the beginning of fermentation (JP0), the metabolic pathway of

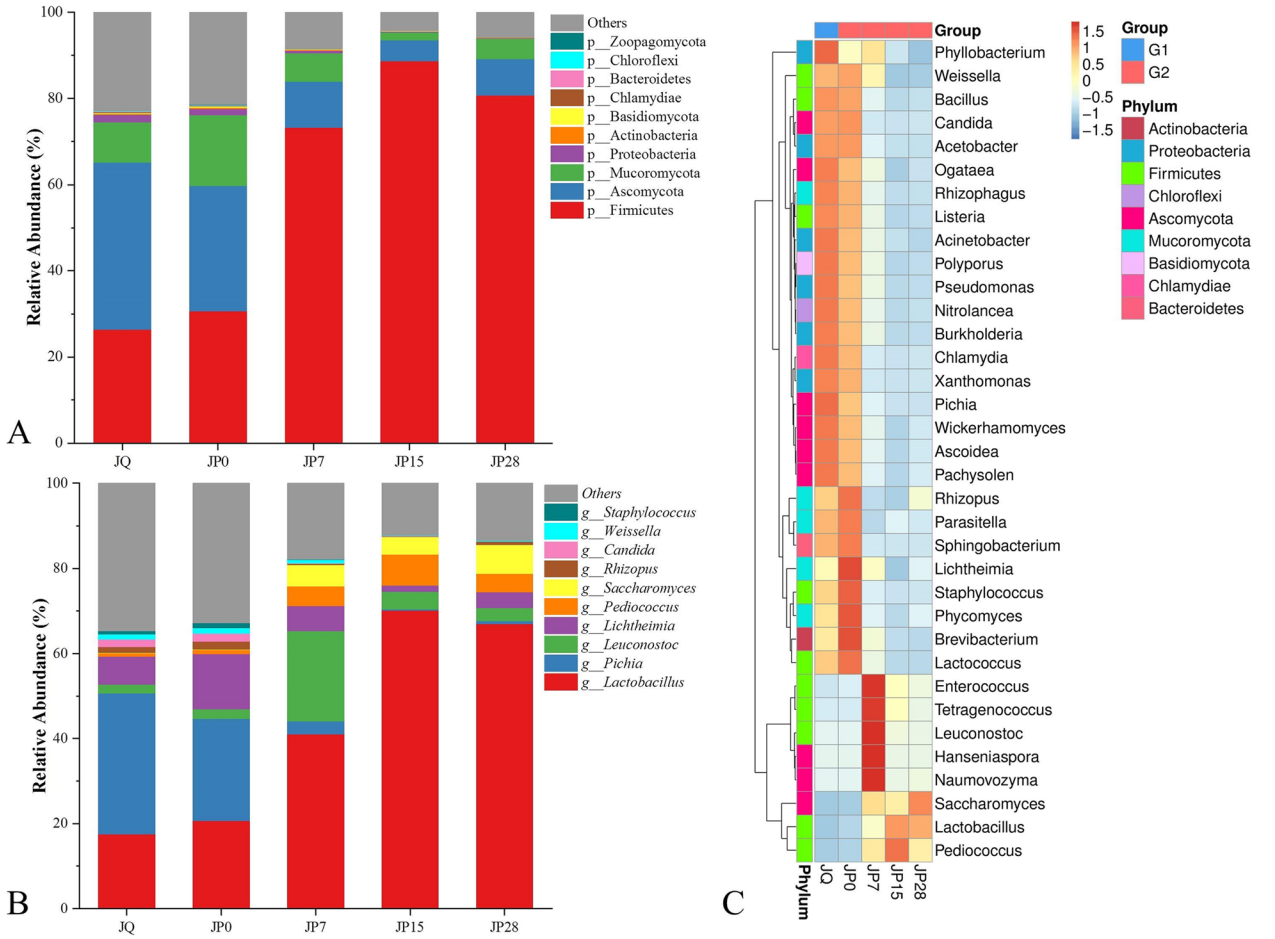

Fig. 6 Microbial community diversity at phylum (**A**) and genus level (**B**) based on metagenomic sequencing (top 10 in relative abundance). Abundance clustering heat of microbial (top 35) map genus level (**C**)

microorganisms was mainly oxidative phosphorylation (ko00190). The primary metabolic function during fermentation gradually evolved into carbohydrate metabolism (purine metabolism (ko00230), pyrimidine metabolism (ko00240), pyruvate metabolism (ko00620), starch and sucrose metabolism (ko00500), glycolysis/gluconeogenesis (ko00010), amino sugar and nucleotide sugar metabolism (ko00520), pentose phosphate pathway (ko00030), galactose metabolism (ko00052)), amino acid metabolism (alanine, aspartate, and glutamate metabolism (ko00250), cysteine and methionine metabolism (ko00270), and nucleotide metabolism (purine metabolism (ko00230), and pyrimidine metabolism (ko00240).

KEGG analysis showed that the microbial communities in fermented grains were mainly growing and reproducing in the early fermentation stage. As the progress of fermentation, microorganisms metabolize carbohydrates and amino acids in raw materials to produce small molecular compounds (such as ethanol, lactic acid, and ethyl acetate).

Functional gene annotation by blasting to CAZy databases

Figure 10 shows the results of functional gene annotation based on the CAZy database. All genes were grouped into six categories at level 1 of CAZy. Glycoside hydrolases (GH) and glycosyl transferases (GT) were annotated in higher numbers (Fig. 10A) and were also the predominant carbohydrate enzymes in all samples (Fig. 10B). Figure 10C shows the clustering results of functional abundance at level 2 of CAZy. The main GHs in JQ are endo-polygalacturonase (GH28), mannosyl-oligosaccharide 1,2-α-mannosidase (GH38), and lactase (GH1), which is similar to that in JP0 samples. As fermentation progresses, lactase (GH1), murein polymerase (GT51), and levansucrase (GH32) increase and become the significant enzymes in JP7. However, in the middle and late stages of fermentation, α, α-trehalase (GH65), lysozyme (GH73, GH25), (1, 4)-α-D-glucan 1-α-D-glucosylmutase (GH13), and mannosylglycoprotein endo-β-mannosidase (GH2) increased and became the main enzymes.

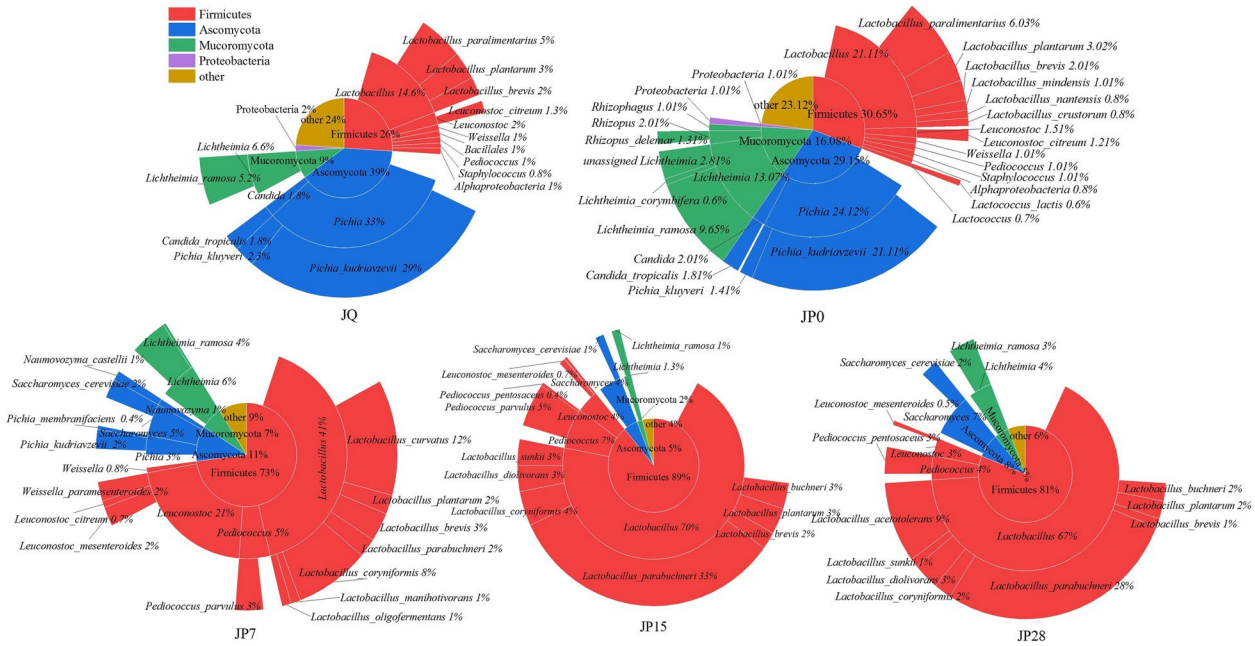

Fig. 7 Microbial community structure in each sample (abundance > 0.3%). The circles from inside to outside are the species abundance at phylum, genus, and species level, and the same color represents the same phylum

In summary, the types of carbohydrate enzymes metabolized by microorganisms constantly change at different fermentation stages. In the early fermentation stage, hydrolysis enzymes are mainly generated to decompress macromolecular carbohydrates. In addition, the abundance of lysozyme increased at the late stage of fermentation, which promoted the dissolution of dead colonies. In addition, the abundance of lysozyme increased at the late stage of fermentation, which promoted the dissolution of dead colonies.

Discussion

The analysis of microbial community structure in fermented grains of *Baijiu* helped understand the rule of microbial succession in the fermentation system, which was essential to control the quality of liquor and was beneficial to the mining of microbial resources. In this study, we use amplicon (16S rDNA and ITS) sequencing and metagenomic sequencing to analyze the diversity and functions of microbial communities in the fermented grains and *Daqu* of *Baijiu*.

By amplicon sequencing, we found that *Lactobacillus* and *Issatchenkia* were identified as the dominant bacteria and fungi during the fermentation process, respectively, and the community diversity of bacteria in fermented grains was higher than that of fungi. However, amplicon sequencing can only obtain the community information of bacteria or fungi by amplifying the 16S rDNA or ITS gene sequence, so we cannot obtain the specific distribution of bacteria or fungi in the entire fermentation system. Metagenomes can compensate for this defect and obtain complete microbial community information.

Metagenome sequencing results showed that *Lactobacillus*, *Weissella*, *Issatchenkia*, *Pichia*, *Candida*, and *Lichtheimia* were the dominant microorganisms in *Daqu*. It is similar to the microbial structure of *Daqu* studied by Yang et al. (2021). Moreover, we found that yeasts were more abundant than LAB at the beginning of fermentation, and the *Lactobacillus*, *Issatchenkia*, or *Pichia* in fermented grains are mainly provided by *Daqu*, while *Saccharomycopsis* may come from raw materials or the environment (Du et al. 2019). With the progress of fermentation,

Lactobacillus and *Weissella* coexist and play an important role in many fermented foods (An et al. 2021; Yongsawas et al. 2022).

In addition, some less abundant microorganisms in *Daqu* and fermented grains, such as *Bacillus*, *Aspergillus*, and *Lichtheimia*, also played essential roles during fermentation. *Aspergillus* and *Lichtheimia* were present in each sample and associated with amino acid metabolism (Liang et al. 2020). Some studies have found that *Bacillus* and yeast are the most effective microorganisms in the first 10 days of *Baijiu* fermentation (Wang & Xu 2019). In this study, *Bacillus* was only present in *Daqu* and the samples on the 28th day of fermentation, and the abundance was low. However, the relative

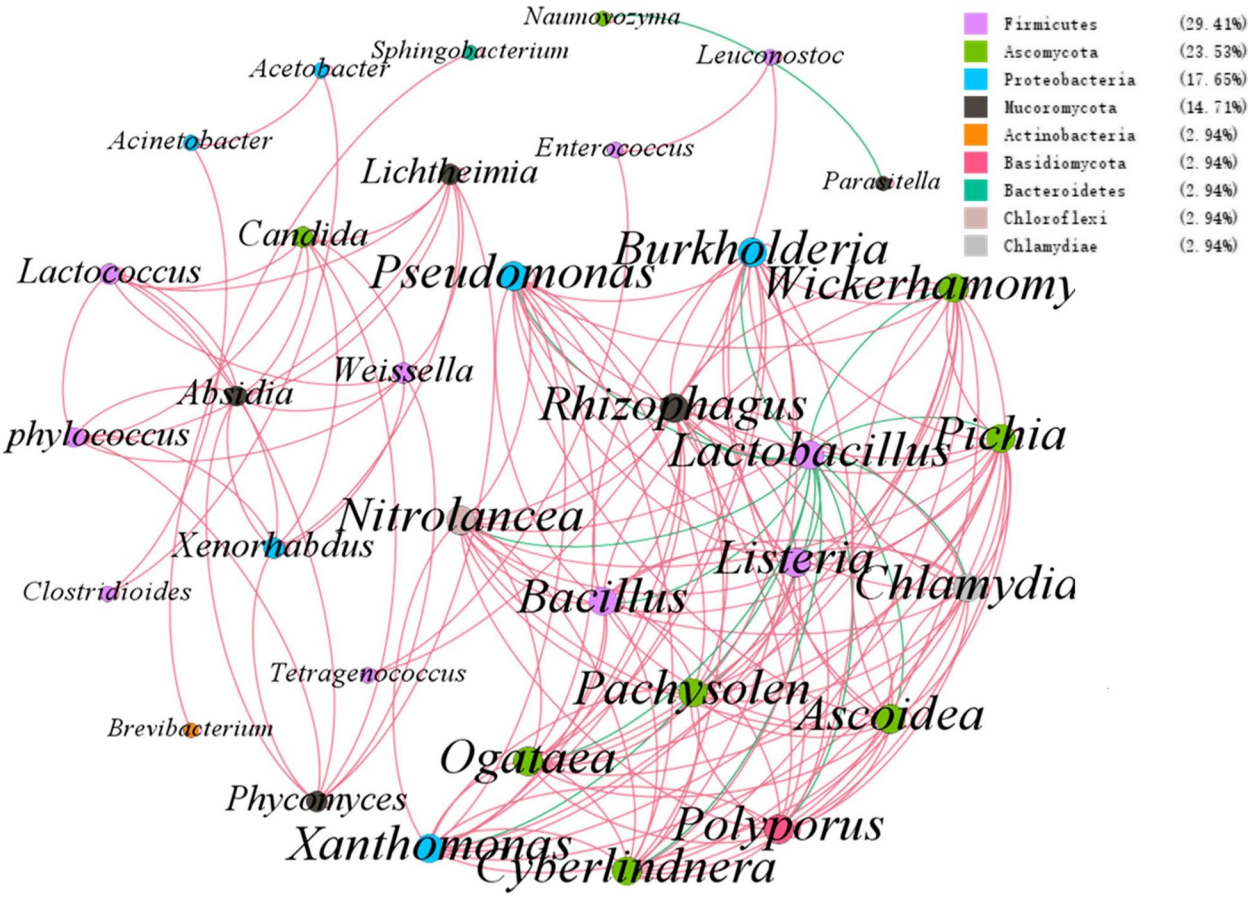

Fig. 8 Correlation network of the microbial community. Different nodes represent different genera, node size represents the degree of connectivity of the genus, the same color represents the level of the same phylum, the thickness of the line represents the size of correlation, the red line indicates positive correlation, and the blue line indicates negative correlation

abundance of Jiang-flavor *Baijiu* was 13.32–31.98% (Wang et al. 2021). When fermented grains' acidity and ethanol content increased (Hu et al. 2021), the growth environment for microorganisms became harsh, increasing the abundance of microorganisms with good tolerance, such as *Bacillus*.

Amplicon sequencing can only identify microorganisms at the genus level but cannot obtain specific information about species at the species level. Metagenomic sequencing can sequence and analyze the entire genome of a sample, which can obtain more detailed information on microbial classification and genes. Annotation of species at the species level using metagenomic sequencing showed that

Pichia kudriavzevii, *Lichtheimia ramosa*, *Companilactobacillus paralimentarius*, *Lentilactobacillus parabuchneri*, and *Lactobacillus acetotolerans* were the main strains in the fermentation process.

In addition, metagenomic sequencing can also obtain the functional information of microbial communities, which can better obtain the information of microorganisms in a sample. In this study,

KEGG and CAZy annotations revealed the metabolic functions of microorganisms in fermented grains. The annotated functions of JQ and JP0 samples were similar, and energy metabolism was the primary function. Carbohydrate metabolism and amino acid metabolism become the main functions in fermentation's middle and late stages. Glycoside hydrolases (GH) and glycosyl transferases (GT) are the major enzymes produced by microorganisms. The results were similar to those of Zhang et al. (Zhang et al. 2022). Studies have found that ethanol production in fermented grains is mainly due to yeasts participating in the glycolytic pathway to promote glycation, lipid production, and alcohol fermentation (Huang et al. 2017a, b; Li et al. 2018; Q et al. 2012; Xu et al. 2022). Aroma components are associated with amino acid metabolism and biosynthesis (An et al. 2020; Zhao et al. 2016), and the dynamic interaction between yeast and LAB promotes the metabolism of flavor chemicals (Tan et al. 2022).

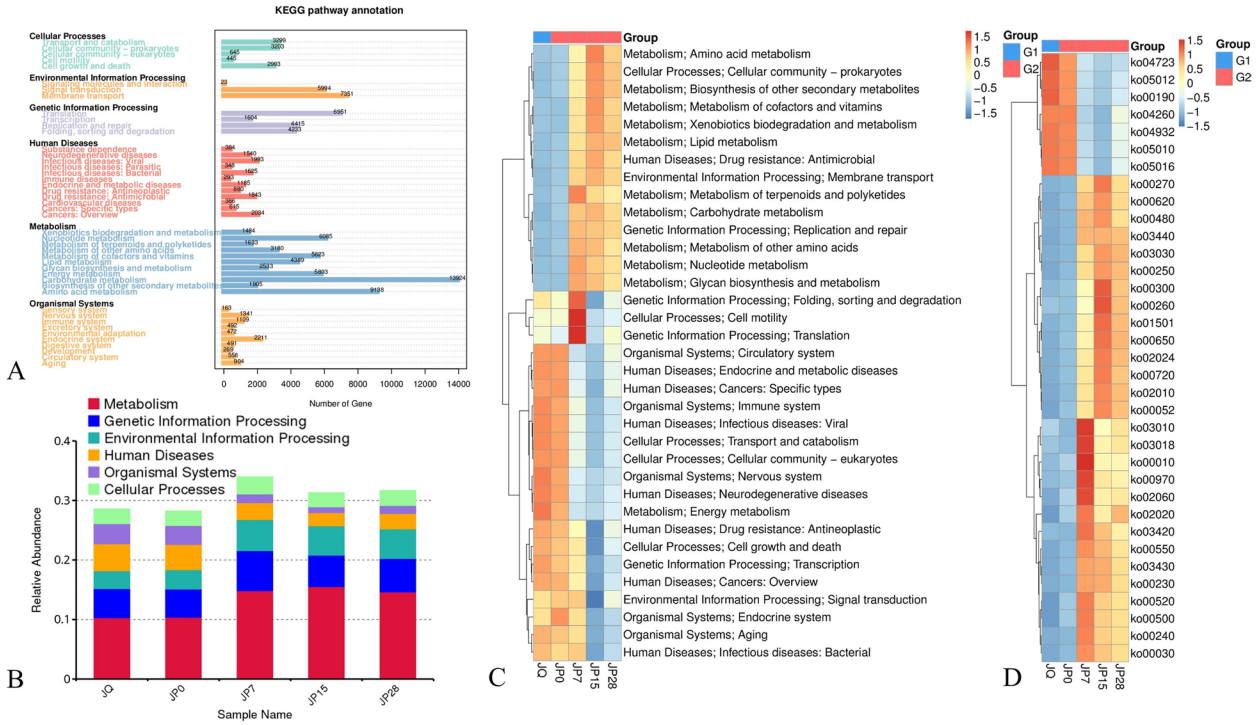

Fig. 9 Functional gene annotation results based on the KEGG database. **A** Statistical chart of the number of KEGG unigenes. The numbers on the bar chart represent the number of unigenes on the annotation. **B** Relative abundance histogram at level 1 of KEGG. **C**, **D** Cluster heat map at levels 2 and 3 of KEGG; the cluster tree on the left is the functional cluster tree. The corresponding value of the middle heat map is the *Z*-value obtained after the standardization of the relative abundance of each line of functions

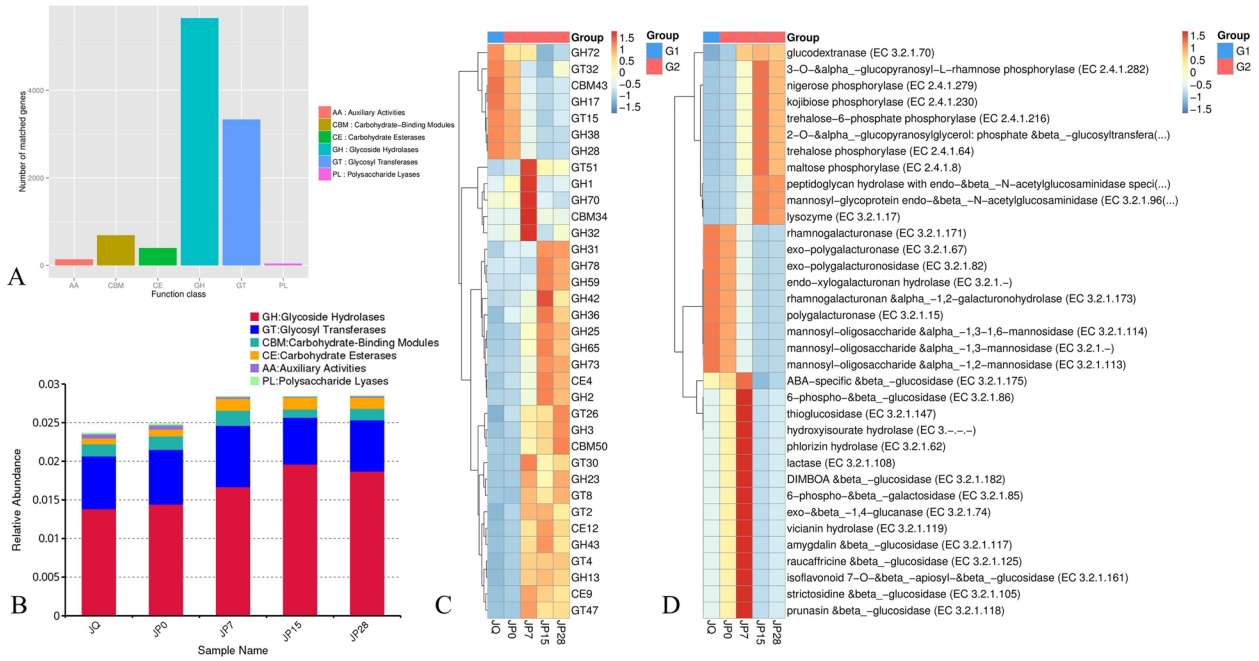

Fig. 10 Functional gene annotation results based on CAZy database. **A** The statistical chart of CAZy unigenes. **B** Relative abundance bar graph at level 1 of CAZy. **C**, **D** Cluster heat map at level 2 and EC of CAZy

It is worth noting that the difference between the two sequencing methods was mainly reflected in Issatchenkia and Pichia, which were closely related. Furthermore, the difference was mainly due to the different databases used. For amplicon species annotation, the SILVA database was used for bacteria and UNITE database for fungi. Metagenomic sequencing is an NR database, and the species information obtained by annotation is more comprehensive, including bacteria, fungi, archaea, and viruses. By comparing the two sequencing methods, we suggest that metagenomic sequencing should be used in the microbial studies of traditional fermented foods to understand the microbial fermentation mechanism better.

Conclusions

Amplicon sequencing can better analyze microbial diversity but can only detect species information at the genus level. In contrast, metagenome sequencing can explore the microbial structure and obtain functional information about microorganisms. This study combined amplicon sequencing and metagenomic sequencing technology to reveal the microbial community structure and gene function prediction during the fermentation of Fen-flavor *Baijiu*. The Spearman coefficient was used to analyze the correlation between microorganisms. The results showed that at the beginning of fermentation, the microorganisms were mainly from *Daqu*, and the abundance of yeast was higher than that of LAB. With the progress of fermentation, the abundance of yeast decreased, while lactic acid bacteria continued to increase. At the later fermentation stage, the microbial community structure tended to be stable, LAB became the main dominant species, and the relative abundance of yeast increased slightly. Functionally, fermented grains' crucial functions were amino acid and carbohydrate metabolism. Finally, we concluded that fermented grains could be a good strain pool for development. In the future, the physicochemical properties and aroma components in fermented grains could be determined, and the correlation between physical and chemical composition and microorganisms could be analyzed further to clarify the specific functions of microorganisms and explore the resources of microbial strains, which is also one of the research hotspots in *Baijiu*.

Acknowledgements
We gratefully acknowledge the support of the Shanxi Xinghuacun Fen Distillery Co., Ltd. (Contract No.: 01130121110114, 01130122090116).

Authors' contributions
Conceptualization, TX and JZ; methodology, JZ; software, TB; validation, JZ and TW; formal analysis, TX and JC; investigation, TB and SF; resources, SF; data curation, TB; writing—original draft preparation, TX; writing—review and editing, JZ, TB, and SF; visualization, TX and ZH; supervision, SF; project administration, SF; funding acquisition, JZ, BB, TB, and SF. All authors have read and agreed to the published version of the manuscript.

Funding
This research was supported by General Youth Fund Project of Shanxi Province (20210302124511) and Xinghuacun College Open Project Foundation (XCSXU-KF-202201, XCSXU-KF-202203, XCSXU-KF-202208, XCSXU-KF-202211).

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All listed authors consented to the submission of this manuscript for publication.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]College of Life Sciences, Shanxi University, Taiyuan 030006, China. [2]Shanxi Key Laboratory of Research and Utilization of Characteristic Plant Resources, Shanxi University, Taiyuan 030006, China. [3]Taiyuan Market Supervision Comprehensive Administrative Law Enforcement Team, Taiyuan 030024, China. [4]Shanxi Food Research Institute (Co., Ltd.), Taiyuan 030024, China. [5]Institute of Biotechnology, Key Laboratory of Chemical Biology and Molecular Engineering, Ministry of Education, Shanxi University, Taiyuan 030006, China.

Received: 11 January 2023 Accepted: 5 April 2023
Published online: 22 May 2023

References

An F, Li M, Zhao Y et al (2020) Metatranscriptome-based investigation of flavor-producing core microbiota in different fermentation stages of dajiang, a traditional fermented soybean paste of Northeast China. Food Chem 343:128509

An F, Sun H, Wu J et al (2021) Investigating the core microbiota and its influencing factors in traditional Chinese pickles. Food Res Int 147:110543

Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 11(12):2639–2643

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP (2016) DADA2: high-resolution sample inference from Illumina amplicon data. Nature Methods 13(7):581–3

Chakravorty S, Helb D, Burday M, Connell N, Alland D (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods 69(2):330–339. https://doi.org/10.1016/j.mimet.2007.02.005

Davis CD (2016) The gut microbiome and its role in obesity. Nutr Today 51(4):167–174. https://doi.org/10.1097/NT.0000000000000167

Du H, Wang X, Zhang Y, Xu Y (2019) Exploring the impacts of raw materials and environments on the microbiota in Chinese Daqu starter. Int J Food Microbiol 297:32–40

He GQ, Liu TJ, Sadiq FA, Gu JS, Zhang GH (2017) Insights into the microbial diversity and community dynamics of Chinese traditional fermented foods from using high-throughput sequencing approaches. J Zhejiang Univ Sci B 18(4):289–302. https://doi.org/10.1631/jzus.B1600148

Hong J, Tian W, Zhao D (2020) Research progress of trace components in sesame-aroma type of baijiu. Food Res Int 137:109695. https://doi.org/10.1016/j.foodres.2020.109695

Hou Q, Wang Y, Cai W, et al (2022) Metagenomic and physicochemical analyses reveal microbial community and functional differences between three types of low-temperature Daqu. Food Res Int 156:111167. https://doi.org/10.1016/j.foodres.2022.111167

Hu X, Tian R, Wang K, Cao Z, Yan P, Li F, He P (2021) The prokaryotic community, physicochemical properties and flavors dynamics and their correlations

in fermented grains for Chinese strong-flavor Baijiu production. Food Res Int 148:110626. https://doi.org/10.1016/j.foodres.2021.110626
Huang X, Fan Y, Lu T, Kang J, Pang X, Han B, Chen J (2020b) Composition and metabolic functions of the microbiome in fermented grain during light-flavor Baijiu fermentation. Microorganisms 8(9):1281. https://doi.org/10.3390/microorganisms8091281
Huang, Y, Huang Y, Huang Y et al (2017a) Metatranscriptomics reveals the functions and enzyme profiles of the microbial community in Chinese Nong-flavor liquor starter. Front Microbiol 8:01747
Huang Y, Yi Z, Jin Y, Zhao Y, He K, Liu D, ... Zhao H (2017) New microbial resource: microbial diversity, function and dynamics in Chinese liquor starter. Sci Rep 7(1):14577. https://doi.org/10.1038/s41598-017-14968-8
Huang F, Pan L, He Z, Zhang M, Zhang M (2020) Identification, interactions, nitrogen removal pathways and performances of culturable heterotrophic nitrification-aerobic denitrification bacteria from mariculture water by using cell culture and metagenomics. Sci Total Environ 732. https://doi.org/10.1016/j.scitotenv.2020.139268
Kang J, Hu Y, Jia L, Zhang M, Zhang Z, Huang X, ... Han BZ (2022) Response of microbial community assembly and succession pattern to abiotic factors during the second round of light-flavor Baijiu fermentation. Food Res Int 162(Pt A):111915. https://doi.org/10.1016/j.foodres.2022.111915
Kang J, Zheng X, Yang X, Li H, Cheng J, Fan L, ... Han B (2022) Contrasting summer versus winter dynamic microbial communities and their environmental driving factors in the solid-state saccharification process of Fuyu-flavor Baijiu. Food Res Int 154. https://doi.org/10.1016/j.foodres.2022.111008
Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Dorrestein PC (2018) Best practices for analysing microbiomes. Nat Rev Microbiol 16(7):410–422. https://doi.org/10.1038/s41579-018-0029-9
Kong Y, Wu Q, Zhang Y, Xu Y (2014) In situ analysis of metabolic characteristics reveals the key yeast in the spontaneous and solid-state fermentation process of Chinese light-style liquor. Appl Environ Microbiol 80(12):3667–3676. https://doi.org/10.1128/AEM.04219-13
Kruskal JB (1964) Nonmetric multidimensional scaling: a numerical method. Psychometrika 29(2):115–129
Li S, Deng Y, Wang Z et al (2020) Exploring the accuracy of amplicon-based internal transcribed spacer markers for a fungal community. Mol Ecol Resour 20(1):170–184
Li W, Chen SJ, Wang JH, Zhang CY, Shi Y, Guo XW, ... Xiao DG (2018) Genetic engineering to alter carbon flux for various higher alcohol productions by Saccharomyces cerevisiae for Chinese Baijiu fermentation. Appl Microbiol Biotechnol 102(4):1783–1795. https://doi.org/10.1007/s00253-017-8715-5
Liang Z, Lin X, He Z, Su H, Li W, Ren X (2020) Amino acid and microbial community dynamics during the fermentation of Hong Qu glutinous rice wine. Food Microbiol 90. https://doi.org/10.1016/j.fm.2020.103467
Liu C, Feng S, Wu Q, Huang H, Chen Z, Li S, Xu Y (2019a) Raw material regulates flavor formation via driving microbiota in Chinese liquor fermentation. Front Microbiol 10:1520. https://doi.org/10.3389/fmicb.2019.01520
Liu Z, Li J, Wei B et al (2019b) Bacterial community and composition in Jiang-shui and Suan-cai revealed by high-throughput sequencing of 16S rRNA. Int J Food Microbiol 306(C):108271
Ma S, Luo H, Zhao D, Qiao Z, Zheng J, An M, Huang D (2022) Environmental factors and interactions among microorganisms drive microbial community succession during fermentation of Nongxiangxing Daqu. Bioresour Technol 345. https://doi.org/10.1016/j.biortech.2021.126549
Pang XN, Huang XN, Chen JY, Yu HX, Wang XY, Han BZ (2020) Exploring the diversity and role of microbiota during material pretreatment of light-flavor Baijiu. Food Microbiol 91. https://doi.org/10.1016/j.fm.2020.103514
Plaza-Diaz J, Ruiz-Ojeda FJ, Gil-Campos M, Gil A (2019) Mechanisms of action of probiotics. Adv Nutr 10(suppl_1):S49–S66. https://doi.org/10.1093/advances/nmy063
Porto M. C. W, Kuniyoshi T. M, Azevedo P. O. S, Vitolo M, Oliveira R. P. S (2017) Pediococcus spp.: an important genus of lactic acid bacteria and pediocin producers. Biotechnol Adv 35(3):361–374. https://doi.org/10.1016/j.biotechadv.2017.03.004
Sala-Comorera L, Caudet-Segarra L, Galofre B, Lucena F, Blanch AR, Garcia-Aljaro C (2020) Unravelling the composition of tap and mineral water microbiota: divergences between next-generation sequencing techniques and culture-based methods. Int J Food Microbiol 334. https://doi.org/10.1016/j.ijfoodmicro.2020.108850
Tan Y, Du H, Zhang H et al (2022) Geographically associated fungus-bacterium interactions contribute to the formation of geography-dependent flavor during high-complexity spontaneous fermentation. Microbiol Spectr 10(5):e01844–22
Wang B, Wu Q, Xu Y et al (2020) Synergistic effect of multiple saccharifying enzymes on alcoholic fermentation for Chinese Baijiu production. Appl Environ Microbiol 86(8):13–20
Wang L (2022) Research trends in Jiang-flavor baijiu fermentation: from fermentation microecology to environmental ecology. J Food Sci 87(4):1362–1374. https://doi.org/10.1111/1750-3841.16092
Wang H, Xu Y (2019) Microbial succession and metabolite changes during the fermentation of Chinese light aroma-style liquor. J Inst Brew 125(1):162–170. https://doi.org/10.1002/jib.544
Wang Q, Wang C, Xiang X, Xu H, Han G (2022) Analysis of microbial diversity and succession during Xiaoqu Baijiu fermentation using high-throughput sequencing technology. Eng Life Sci 22(7):495–504. https://doi.org/10.1002/elsc.202200015
Wang, H., Huang, Y., & Huang, Y. (2021). Microbiome diversity and evolution in stacking fermentation during different rounds of Jiang-flavoured Baijiu brewing. Lwt, 143. https://doi.org/10.1016/j.lwt.2021.111119
Wensel, C. R., Pluznick, J. L., Salzberg, S. L., & Sears, C. L. (2022). Next-generation sequencing: insights to advance clinical investigations of the microbiome. J Clin Invest, 132(7). https://doi.org/10.1172/JCI154944
Wu Q, Xu Y, Chen L (2012) Diversity of yeast species during fermentative process contributing to Chinese Maotai-flavour liquor making. Lett Appl Microbiol 55(4):301–7
Wu Q, Zhu Y, Fang C, Wijffels RH, Xu Y (2021) Can we control microbiota in spontaneous food fermentation? – Chinese liquor as a case example. Trends Food Sci Technol 110:321–331. https://doi.org/10.1016/j.tifs.2021.02.011
Xu Y, Sun B, Fan G, Teng C, Xiong K, Zhu Y, ... Li X (2017) The brewing process and microbial diversity of strong flavour Chinese spirits: a review. Journal of the Institute of Brewing 123(1):5–12. https://doi.org/10.1002/jib.404
Xu J, Sun L, Xing X, Sun Z, Gu H, Lu X, ... Ren Q (2020) Culturing bacteria from fermentation pit muds of Baijiu with culturomics and amplicon-based metagenomic approaches. Front Microbiol 11:1223. https://doi.org/10.3389/fmicb.2020.01223
Xu Y, Zhao J, Liu X, Zhang C, Zhao Z, Li X, Sun B (2022) Flavor mystery of Chinese traditional fermented Baijiu: the great contribution of ester compounds. Food Chem 369. https://doi.org/10.1016/j.foodchem.2021.130920
Yang Y, Wang ST, Lu ZM, Zhang XJ, Chai LJ, Shen CH, Xu ZH (2021) Metagenomics unveils microbial roles involved in metabolic network of flavor development in medium-temperature daqu starter. Food Res Int 140. https://doi.org/10.1016/j.foodres.2020.110037
Yongsawas, R., Inta, A., Kampuansai, J., Pandith, H., Suwannarach, N., Lamyong, S., … Disayathanoowat, T. (2022). Bacterial communities in Lanna Phak-Gard-Dong (Pickled mustard green) from three different ethnolinguistic groups in Northern Thailand. Biology (Basel), 11(1). https://doi.org/10.3390/biology11010150
Zhang J, Liu S, Sun H, Jiang Z, Xu Y, Mao J, ... Mao J (2022) Metagenomics-based insights into the microbial community profiling and flavor development potentiality of Baijiu Daqu and Huangjiu wheat Qu. Food Res Int 152. https://doi.org/10.1016/j.foodres.2021.110707.
Zhao CJ, Schieber A, Ganzle MG (2016) Formation of taste-active amino acids, amino acid derivatives and peptides in food fermentations - a review. Food Res Int 89(Pt 1):39–47. https://doi.org/10.1016/j.foodres.2016.08.042
Zhao CC, Eun JB (2020) Shotgun metagenomics approach reveals the bacterial community and metabolic pathways in commercial hongeo product, a traditional Korean fermented skate product. Food Res Int 131. https://doi.org/10.1016/j.foodres.2020.109030
Zheng X, Yan Z, Han et al (2012) Complex microbiota of a Chinese " Fen " liquor fermentation starter (Fen - Daqu), revealed by culture-dependent and culture-independent methods. Food Microbiol 31(2):293–300
Zuo W, Wang B, Bai X, Luan Y, Fan Y, Michail S, Sun F (2022) 16S rRNA and metagenomic shotgun sequencing data revealed consistent patterns of gut microbiome signature in pediatric ulcerative colitis. Sci Rep 12(1):6421. https://doi.org/10.1038/s41598-022-07995-7

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.