World Journal of Microbiology and Biotechnology (2023) 39:337
https://doi.org/10.1007/s11274-023-03772-4

RESEARCH

Ultra-high-depth macrogenomic sequencing revealed differences in microbial composition and function between high temperature and medium–high temperature *Daqu*

Yurong Wang[1,2] · Jianshe Gai[3] · Qiangchuan Hou[1,2] · Huijun Zhao[1,2] · Chunhui Shan[4] · Zhuang Guo[1,2,3]

Received: 22 May 2023 / Accepted: 18 September 2023 / Published online: 10 October 2023

Abstract
Complex microorganisms in *Daqu* of different temperatures play a vital role in the taste, flavor and quality of *Baijiu* during fermentation. However, understanding the functional diversity of the whole microbial community between the *Daqus* of two different temperatures (high temperature *Daqu*, HD and medium–high temperature *Daqu*, MD) remains a major challenge. Here, a systematic study of the microbial diversity, functions as well as physiological and biochemical indexes of *Daqu* are described. The results revealed that the *Daqu* exhibited unique characteristics. In particular, the diversity of microorganisms in HD and MD was high, with 44 species including 14 novel species (*Sphingomonas* sp. is the main novel species) detected in all samples. Their profiles of carbohydrate-active enzymes and specific functional components supported the fact that these species were involved in flavor formation. The *Daqu* microbiome consisted of a high proportion of phage, providing evidence of phage infection/genome integration and horizontal gene transfer from phage to bacteria. Such processes would also regulate *Daqu* microbiomes and thus flavor quality. These results enrich current knowledge of *Daqu* and can be used to promote the development of *Baijiu* fermentation technology.

Keywords Different temperature · Metagenomics · Bionic technology · *Daqu* · Diversity

Introduction

Baijiu, also known as soju, is mainly produced in areas such as mainland China and Taiwan (Xu et al. 2017), and being a traditionally distilled spirit, its brewing process has been practiced for over a thousand years in China (Jin et al. 2017). Based on aroma style, Chinese *Baijiu* can be divided into Maotai-flavored *Baijiu*, strong fragrance *Baijiu*, clear fragrance *Baijiu* and phoenix fragrance *Baijiu*, just to name a few (Hong et al. 2020). Of these, the sauce fragrance style is among the most typical ones, and it is loved by most consumers (Wang et al. 2019). In fact, compared with other types of liquor, the soy sauce one has a more complex brewing process, with the high-temperature stacking step being the most unique one during its production (Hao et al. 2021). This is because the high-temperature accumulation process not only enriches and proliferates microorganisms in the production environment (Li et al. 2022), but also produces aromatic substances such as alcohols, esters, organic acids as well as other precursors of aromatic substances as a result of microbial metabolism. Altogether, these processes provide raw materials for good fermentation, thus giving soy sauce-spiced liquor its special aroma (Wang et al. 2019). So far, a number of studies have shown that the quality of Maotai-flavored liquor is related to many factors, including the type, the raw materials, the water source and the production method of *Daqu* (Hao et al. 2021; Hou et al. 2022). In particular, the type of *Daqu* can even directly determine the final quality of Maotai-flavor liquor (Zheng et al. 2011). FAN's research also proves that the main role of the *Daqu* is

✉ Zhuang Guo
guozhuang@vip.163.com

1 Hubei Provincial Engineering and Technology Research Center for Food Ingredients, Hubei University of Arts and Science, Xiangyang, Hubei Province, People's Republic of China

2 Xiangyang Lactic Acid Bacteria Biotechnology and Engineering Key Laboratory, Hubei University of Arts and Science, Xiangyang, Hubei, People's Republic of China

3 Xinjiang Sishi Avenue Wine Co., Ltd, Huyanghe, Xinjiang Autonomous Region, People's Republic of China

4 School of Food Science, Shihezi University, Shihezi, Xinjiang Autonomous Region, People's Republic of China

to provide the microbial flora and enzymes required for the fermentation system of *Baijiu* and to make a special contribution to the flavor (Fan et al. 2018). *Daqu* can be divided into high temperature and medium–high temperature ones according to its fermentation temperature, even though the same raw materials, pretreatment conditions and culture environment are used during the production process. During *Daqu*'s production, temperature greatly influences microorganisms, with different temperatures directly affecting the microflora and enzymes in *Daqu* (Zheng et al. 2011). However, the effects of temperature on the structure and function of microbial communities in *Daqu* are not clear.

In previous studies, researchers have mostly used pure culture methods to resolve microbial composition and diversity in ecological environments (Dong et al. 2020). Over the past decade, with the rapid development of sequencing technology and bioinformatics, researchers have, to some extent, been able to apply amplicon sequencing to resolve the microorganisms present during the fermentation process of *Baijiu* (Chen et al. 2020). However, since amplicon sequencing requires DNA amplification, this step further amplifies systematic errors and distorts the microbial community structure in ecological samples. Hence, with the decreasing cost of sequencing in recent years, macrogenome sequencing technology has gradually been applied as an alternative to parse traditional fermentation systems (Yang et al. 2021). In this context, metagenomic sequencing and multiplex analysis can provide an important means for uncovering the composition and structure analysis of traditional fermented food, including *Daqu* (Hou et al. 2022). The community structure of microorganisms in *Daqu* significantly influences the flavor and quality of *Baijiu*. At the same time, the final quality of *Baijiu* is highly dependent on the dynamic interactions between the microorganisms, the raw materials and the environment in *Daqu*. Ultra-high-depth macro-genome sequencing, combined with bioinformatics analysis, can not only analyze the species diversity and function at the genomic level, but also observe the viral sequences as well as horizontal gene transfer in the environment. It is worth noting that bacteriophage infection of bacteria involved in fermentation is also a common problem that leads to fermentation failure. Hence, it is necessary to analyze the microbial community structure of *Daqu* at different temperatures by applying multiple analytical technology.

Although the ultra-high-depth macro-genome sequencing technology can accurately analyze the microorganisms in environmental samples, yet it is unable to evaluate the quality of *Daqu*, an important factor that allows the effect of microorganisms on the quality of *Daqu* to be evaluated. In this case, electronic tongues and noses represent bionic techniques which can be used to digitally evaluate the sensory quality of *Daqu* samples as a whole, with this approach being widely used in recent years to analyze the sensory quality of traditional fermentation products. Therefore, these features justify the application of the above techniques to this study to further analyze the quality of *Daqu* through multiple analytical platforms.

This study assumes that moderate and high temperature *Daqu* have different sensory qualities which, in turn, are closely related to the bacterial and bacteriophage composition in *Daqu*. For the current work, a total of 10 *Daqu* samples (including 5 *Daqu* at medium–high temperature and 5 *Daqu* at high temperature) were collected from a winery in Hubei, China. Ultra-high-depth macrogenomic sequencing as well as electronic tongue and nose techniques were then used to analyze the relationship between microorganisms and *Daqu* quality. In addition, specific microbial characteristics in *Daqu*, including carbohydrates, secondary metabolites and resistance genes, were further identified. To analyze the important role of *Daqu* in the fermentation process of liquor.

Materials and methods

Sample collection

In this study, a total of 10 *Daqu* samples at the end of fermentation (5 high temperature *Daqu* (numbered HD1–HD5) and 5 medium–high temperature ones (numbered MD1–MD5)) were after which they were collected and sealed in a sterile bag. The samples were then transferred to the laboratory where they were stored at − 80 °C for backup.

Analysis of microbial diversity in *Daqu*

DNA extraction

For the 10 samples of *Daqu*, about 5 g of powder was weighed before using about 1 g for DNA extraction with the DNasy PowerFood Kit. The purity, concentration and integrity of the extracted microbial DNA were tested using 1% agarose gel electrophoresis as well as micro-ultraviolet spectrophotometer (Yang et al. 2020). DNA samples of the required quality were then stored at − 80 °C until required for use.

Shotgun metagenomic sequencing and quality control

For each sample, sequencing libraries were prepared with 1 μg of DNA by using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) according to the manufacturer's recommendations. In this case, indexes were also added to attribute sequences to their respective samples. Shotgun sequencing was eventually performed on a HiSeq

2500 (Illumina, California, USA) platform to generate paired-end reads (151 bp in length) (Yang et al. 2021).

The reads were trimmed using KneadData (Anyaso-Samuel et al. 2021) before aligning them to the barley genome to remove host DNA sequences. For the remaining reads, human as well as low-quality sequences were also removed to yield a total of 293.75 Gb of high-quality sequence data for all samples (average of 229.38 Gb).

Bioinformatic analysis

Megahit (ver. 1.2.9) was used to assemble the high quality reads obtained after quality control into contigs (Li et al. 2015a), with the parameters -k 33,55,77,99,111 -meta, and QUAST (ver. 5.0.0) used to evaluate and screen the assembly results (Gurevich et al. 2013). Vamb (ver. 3.0.3) (Nissen et al. 2021), MetaBAT2 (ver. 2.12.1) (Kang et al. 2019) and Maxbin2 (ver. 2.0) (Maguire et al. 2020) were then used to perform binnings on high quality contigs to obtain the genome of the species (scaffold length threshold of 1500 bp). Finally, MetaPhlan3 (ver. 3.0) (Nousias and Montesanto 2021) allowed the microbial communities in *Daqu* to be annotated based on Bowtie2 (ver. 2.2.9) (Langdon 2015).

The chromaticity of each scaffold was calculated by Bowtie2 (ver. 2.2.9) and SamTools (ver. 1.9) (Li et al. 2009), with the coverage depth of each genome inferred based on the jgi_summarize_bam_contig_depths function. The genomes assembled by the three strategies were combined and evaluated with the Das tool (v 1.1.2) to obtain high quality genomes (Sieber et al. 2018). Finally, metagenome-assembly genomes (MAGs) that met the quality requirements (completeness > 80% and contamination < 10%) were selected. High-quality MAGs were clustered at species-level genome bin (SGB) using dRep (ver. 2.2.4) (Olm et al. 2017) before applying the BLASTn software against the National Center for Biotechnology Information (NCBI) Non-redundant Nucleotide Sequence Database (NT) to annotate the taxonomic status of each MAG (a threshold of 95% identity over 70% of coverage). Phylogeny was inferred using the 400 universal PhyloPhlAn markers based on parameters described by Pasolli, with the resulting phylogenetic trees visualized using iTOL (ver. 5.5.1) (Kapli et al. 2020).

Proteins of the contigs were predicted with Prodigal (ver. 2.6.3) (Fang et al. 2020) before using AntiSMAH (ver. 5.0.28) (Blin et al. 2019), HMMER (ver. 3.3.2) (Eddy 2011) and ABRicate (de Man and Limbago 2016) to predict the genome.

Deep-virfinder (v 1.0) and Vibrant (v1.2.1) were also used to identify prophages in the contigs (> 1500 bp) (Liu et al. 2020) before assessing the quality of bacteriophage virus contigs with CheckV.

Nucleotide sequence accession numbers

The sequence dataset has been deposited in the NCBI Sequence Read Archive (SRA) database (Accession number: PRJNA1010526, https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1010526).

Determination of flavor quality

After accurately weighing 20 g of *Daqu* powder into the electronic nose (PEN3, Airsense, Germany) test bottle, the latter was sealed and placed at room temperature at 30 min. The electronic nose probe was then inserted into the test bottle to digitally determine the specific flavor substances in the *Daqu*. In this case, the substances were analyzed based on its 10 metal test electrodes. The electronic nose was automatically cleaned for 90 s prior to each nose test, with the test itself performed for 60 s. During the test, a response value was generated every second, and based on the resulting data, it was found that the response curves for all samples became flat after 30 s. Therefore, in this study, the average response values of 49 s, 50 s and 51 s were selected as the sample test data for follow-up analysis (Wang et al. 2021).

Determination of taste quality

To 25 g of *Daqu* powder that was accurately weighed into a 250 ml beaker, 100 g of distilled water was added and stirred for 15 min. The mixture was then centrifuged at 10,000 r/min for 10 min, with the resulting supernatant poured into the electronic tongue (SA 402B, INSENT, Japan) cup to be tested. Prior to the test, the electronic tongue system first passed the cleaning and self-test to determine the sourness, bitterness, astringent, saltiness and taste of *Daqu*, as well as aftertaste-A, aftertaste-B and abundance. Each sample was measured four times in parallel, with the average value of the last three test results taken as the test value for follow-up analysis (Wang et al. 2020).

Determination of physical and chemical indicators

The acidity, starch content, fermenting power, liquefying power, saccharifying power, esterifying power, ash content, amino acid nitrogen and alcoholic power of *Daqu* at different temperatures were determined according to the light industry standard QB/T 4257–2011 (General methods of analysis for *Daqu*). The protein content at different temperatures was also determined according to the Application manual of Kay's nitrogen determination. In addition, 3 g of *Daqu* powder was weighed into an aluminum foil sample plate to determine its moisture content with a moisture tester, while in the case of water activity, a moisture absorption analyzer

was used after spreading the *Daqu* powder all over the sample pool.

Plate count

The total number of bacteria in *Daqu* samples at different temperatures was determined according to GB 4789.2-2016 (Food microbiology testing—Determination of total colony count). The count for lactic acid bacteria was carried out according to GB 4789.35-2016 (Food microbiology testing—Lactic acid bacteria test), while for molds and yeast counts, the GB 4789.15-2016 (Food microbiology testing—Mold and yeast count) procedure was followed.

Statistical analysis and data visualization

The "vegan (version 2.6–2)" package in R (version 4.2.2) was used for alpha and beta diversity analyses of microorganisms and phages. Statistical differences between indicators and alpha diversity between groups were then assessed with the Wilcoxon rank sum test. Similarly, the "Psych (version 2.3.6)" package was used for Spearman correlation analysis, principal component analysis and principal coordinate analysis. The "ape (version 5.7-1)", "vegan (version 2.6-2)", "philentropy (version 0.7.0)"and "ggdendro (version 0.1.23)" package was used for consistency analysis while other multivariate statistical analyses, including principal component analysis, principal coordinate analysis, and consistency analysis, were completed with R.

Results

Physiological and biochemical indexes in *Daqu*

The total number of Microorganism in *Daqu* samples at different temperatures was determined by plate count. There was no significant difference in Cultivable total colony count between the two groups ($P = 1.00$, Fig. 1a), but Cultivable mold and yeast count and Cultivable lactic acid bacteria in Daqu in MD group were significantly higher than those in HD group ($P = 0.01$, Fig. 1a). Results of physiological and biochemical indexes revealed significantly lower acidity and water activity for HD compared with MD ($P = 0.01$; $P = 0.03$, Fig. 1b), although HD had a significantly higher alcoholic power than the latter ($P = 0.01$, Fig. 1b). Besides, enzymatic activity was detected according to the light industry standard QB/T 4257–2011. The results further revealed a significantly lower saccharifying power for HD ($P = 0.04$, Fig. 4c), along with a significantly higher esterifying power, fermenting and liquefying power compared with MD ($P = 0.01$, Fig. 1c). Regarding the PCA results, There is a significant difference in the taste and flavor of *Daqu*, and the difference in taste is even greater (Fig. 1d, e).

Daqu microbiota composition and diversity

The microbiota profiles of *Daqu* samples at species-level were detected by MetaPhlan3 and consisted of a total of 121 species from 6 phyla and 25 families. Of these, 14 dominant species with a relative abundance of > 1.00% were identified, as shown in Fig. 2a. Overall, the bacterial composition in *Daqu* was as follows: *Lactobacillus brevis*, *Staphylococcus xylosus*, *Bacillus haynesii*, *Weissella paramesenteroides*, *Leuconostoc citreum*, *Alcelaphine gammaherpesvirus*, *Weissella confuse*, *Weissella cibaria*, *Saccharopolyspora rectivirgula*, *Sphingomonas sp FARSPH*, *Pediococcus pentosaceus*, *Lactobacillus plantarum*, *Kroppenstedtia eburnean* and *Lactobacillus crustorum*. In addition, the average relative abundance of bacteria, *Eukaryota* and viruses were 91.70%, 2.45% and 5.85%, respectively. While most *Daqu* samples were dominated by bacteria, a small number were dominated by *Eukaryota* (Fig. 2b). In terms of alpha-diversity analysis, the results revealed significant differences in species diversity and richness, as measured by Shannon index and S index between the HD and MD groups, as well as within the microbial community composition ($P = 0.01$; Fig. 2c, d). Besides, the *Daqu* microbiota was highly diverse across samples, and clear region-based clustering were identified (Fig. 2e). These results suggested that *Daqu* from different temperatures had different levels of microbial diversity.

MAGs profiles reflected differences in *Daqu* quality

Metagenome-assembled genomes (MAGs) represent single genomes derived by metagenomic binning, and a total of 217 MAGs were assembled across all *Daqu* samples. Incomplete bins with high contamination were removed, and this yielded 44 high-quality MAGs (Completeness > 80% and contamination < 10%; Fig. 3a). These genomes were taxonomically assigned to two phyla (21 Firmicutes and 23 Actinobacteria), 8 genera and 13 species, with the ten bacterial species mainly belonging to LAB (Fig. 3a). Besides, a total of 14 suspected new species (7 from each group) were identified (Compared with all genomes in NCBI and UHGG databases, ANI value is less than 95%), with *Sphingomonas sp*. being the main suspected new species.

A total of 14 representative species-level genome bins (SGBs) were also obtained by clustering analysis of the 44 high-quality MAGs. *Daqu* fermentation temperature and flavor formation involve carbohydrate utilization as well as the synthesis of related metabolites. Thus, carbohydrate-active enzymes (CAZymes) and secondary metabolite clusters in the metagenomes were predicted (Fig. 3b–d). A total of 121 CAZymes were detected across all *Daqu* samples (Fig. 3b),

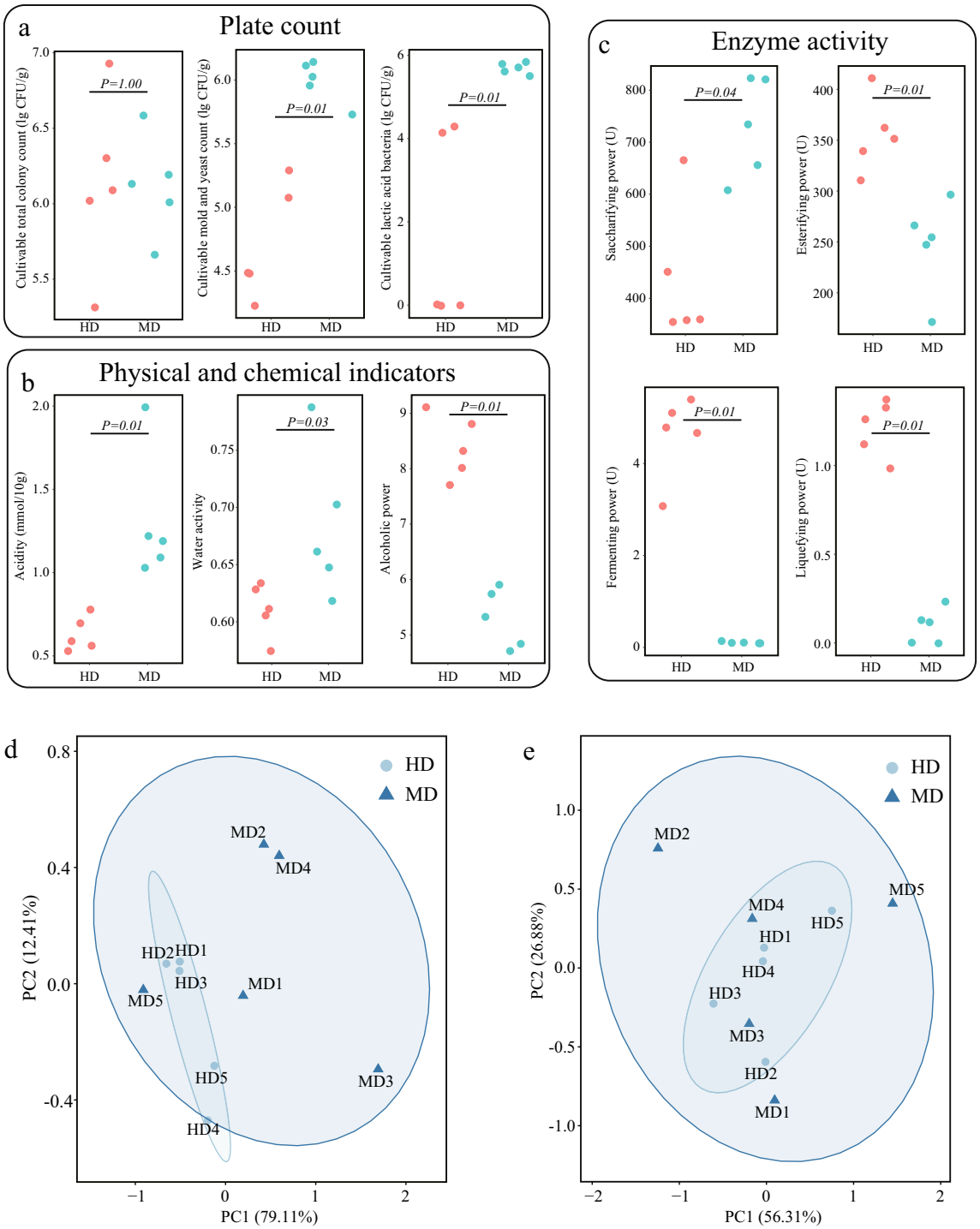


Fig. 1 Physiological and biochemical indexes in *Daqu*. **a** Plate count; **b** Physical and chemical indicators; **c** Enzyme activity; **d** Principal component analysis calculated based on euclidean distance (taste quality); **e** Principal component analysis calculated based on euclidean distance (flavor quality). HD stands for high temperature *Daqu*; MD for medium–high temperature *Daqu*

and their relative abundance revealed significantly higher auxiliary activities (AAs), carbohydrate-binding modules (CBMs), carbohydrate esterases (CEs) and glycoside hydrolases (GHs) in the HD group compared with MD. At the same time, no significant differences were observed in glycosyltransferase (GTs) and polysaccharide lyases (PLs) between the two groups (Fig. 3c). Genes related to 15 secondary metabolite clusters and six carbohydrate metabolism modules were then detected across the 14 SGBs, with the most widely distributed secondary metabolite clusters being

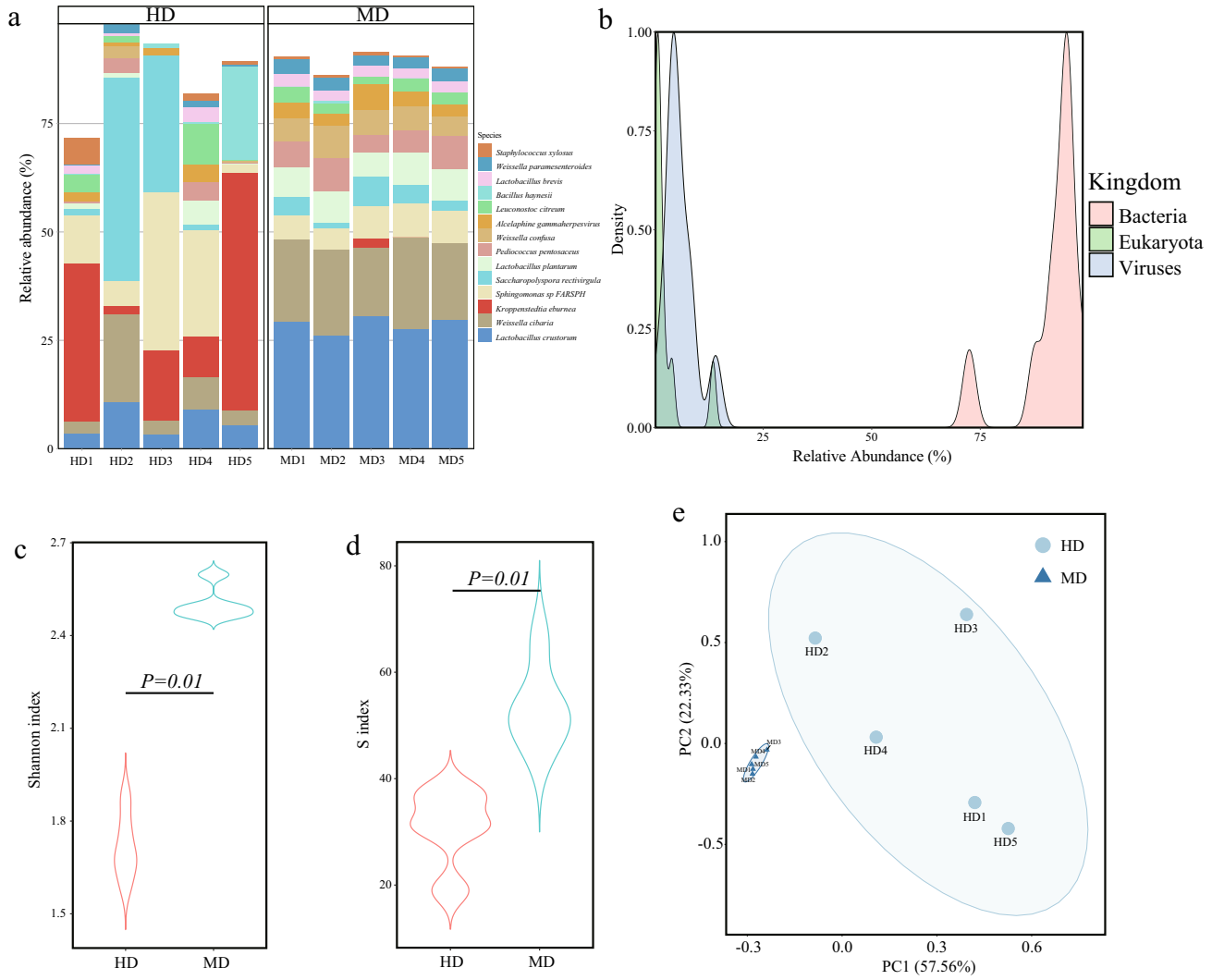


Fig. 2 *Daqu* microbiota composition and diversity. **a** Relative abundance of dominant species that were ≥ 1.0% of the total sequences; **b** Distribution frequency diagram of bacteria, phages and fungi across samples; **c** Boxplots showing values of Shannon index between the two groups; **d** Boxplots showing values of S index between the two groups; **e** Principal coordinate analysis calculated based on Bray–Curtis distance. HD stands for high temperature *Daqu*; MD for medium–high temperature *Daqu*

terpene, type III polyketide synthases (T3PKS), bacteriocin and ectoine (Fig. 3d). In addition, the hits gene kanamycin, macrolide, rifampin, rifamycin, spectinomycin and vancomycin1 were found in HD1-9, HD2-4, HD3-3, HD4-4 and MD4-1 (Fig. 3e).

Prevalence of phages in *Daqu* microbiome

A high proportion of phage sequences was found in *Daqu* microbiota. Considering the important role and significant negative effects of phages in *Daqu*, their sequences and diversity in the samples were analyzed, with the quality of the assembled phage sequences shown in Fig. 4a. Taxonomic profiling showed that the *Daqu* phage sequences mainly belonged to eight families, namely Microviridae, Inoviridae, Tectiviridae, Ackermannviridae, Siphoviridae, Herelleviridae, Podoviridae and Myoviridae (Fig. 4a). In particular, the most abundant phages in this study's dataset belonged to the Siphoviridae family. Overall, HD had a higher but non-significant Shannon index and S index (alpha diversity) compared with MD (Fig. 4b), but significant regional differences was still observed in the phage sequences ($P = 0.02$, Fig. 4c). The PCoA results showed significant separation of phage and *Daqu* bacterial microbiota (Fig. 4d). These results suggested that although the *Daqu* samples contained quite a high proportion of phage sequences, differences in the composition and structure of *Daqu* microbiota between the two compared temperatures could be attributed to the bacterial sequences. The relative abundance of Myoviridae and Podoviridae in the phage microbiome of HD was

significantly higher compared with that of MD. In contrast, the relative abundance of Siphoviridae decreased significantly in the case of HD, compared with MD (Fig. 4e). In terms of correlational analysis, HD5-2 correlated significantly and positively with Podoviridae, while HD2-5 and HD2-6 correlated significantly and negatively with Podoviridae. Finally, HD1-7 correlated significantly and negatively with Herelleviridae (Fig. 4f).

Correlation between SGBs and specific functional components

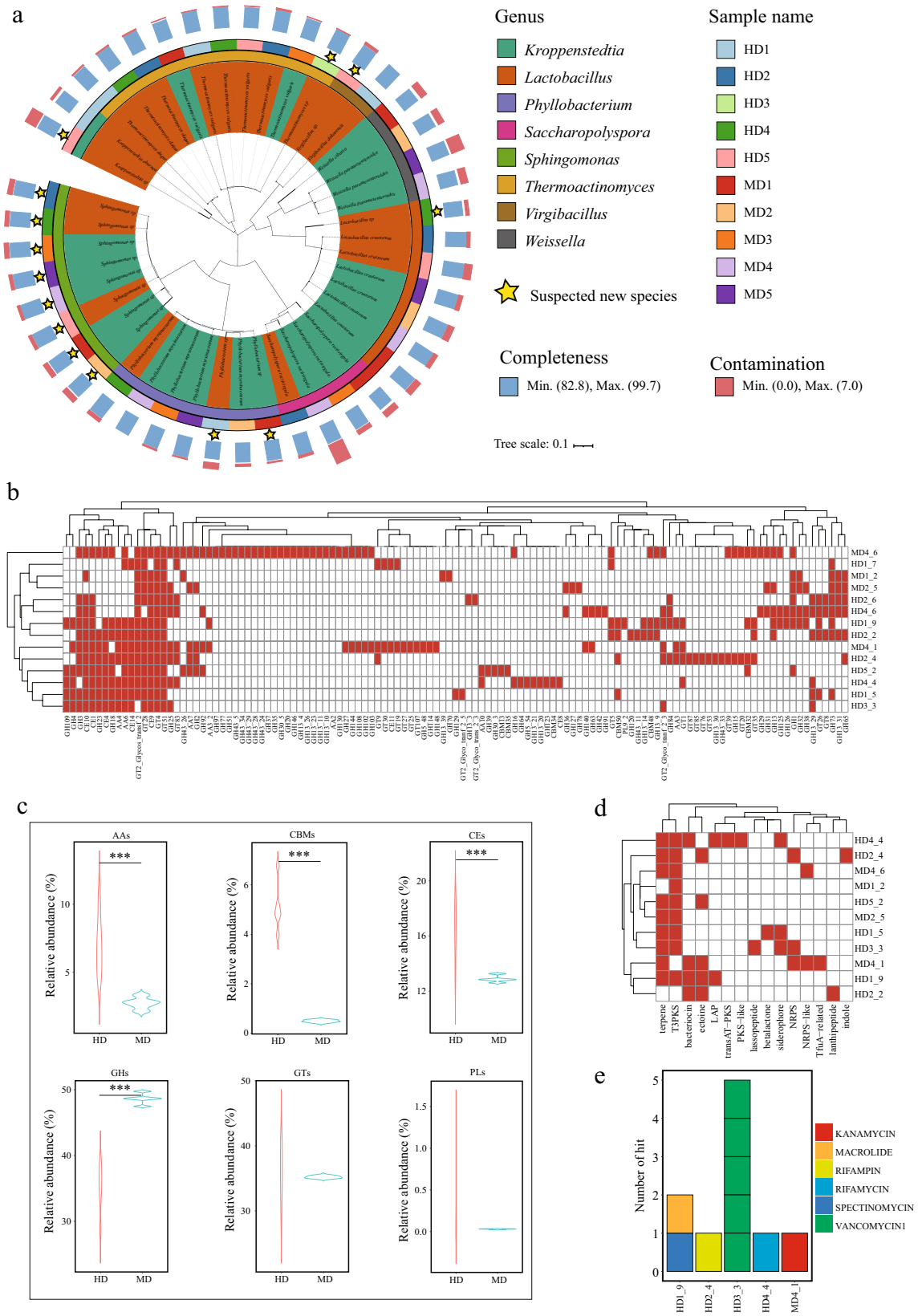
In order to reveal whether the specific functional components that were measured were consistent with the structure and function of *Daqu* at different temperatures, Procrustes analysis was performed. Through this approach, the congruence of two-dimensional shapes, produced from the superimposition of principal component analyses for SGBs with flavor index, enzyme activity and taste index function, were analyzed. The PCA results revealed no significant differences between the SGBs and taste index ($P = 0.4$, Fig. 5a), and as such, it was speculated that differences could exist between the SGBs and specific functional components such as flavor index and enzyme activity ($P = 0.01$; $P = 0.006$, Fig. 5b, c). Thus, Spearman's correlation analysis was performed and, as expected, a significant correlation was found between the microbial community and specific functional components (Spearman's test, $r > 0.6$, $P < 0.05$; Fig. 5d). Moreover, most SGBs showed significant positive correlations with flavor index and liquefaction power. However, few SGBs (HD2-2, HD2-6, HD1-7 and HD4-4) also showed significant negative correlations with saccharifying power (Spearman's test, $r < -0.6$, $P < 0.05$). Furthermore, it is worth noting that HD3-3 was significantly and positively correlated with flavor index (W2W; Spearman's test, $r < 0.6$, $P < 0.05$), but significantly and negatively correlated with saccharifying power (Spearman's test, $r < -0.6$, $P < 0.05$). On the other hand, HD2-2 was significantly and positively correlated with liquefaction power (Spearman's test, $r < 0.6$, $P < 0.05$), but significantly and negatively correlated with saccharifying power (Spearman's test, $r < -0.6$, $P < 0.05$). In the case of HD2-4, the SGB was significantly and positively correlated with taste index (aftertaste-B; Spearman's test, $r < 0.6$, $P < 0.05$), but significantly and negatively correlated with flavor index (W5C; Spearman's test, $r < -0.6$, $P < 0.05$). Finally, HD1-2 and HD1-5 were significantly and positively correlated with liquefaction power (Spearman's test, $r < 0.6$, $P < 0.05$). These results suggested that the different temperatures induce different characteristics in the *Daqu*, and it is worth exploring the functional capacity of the novel species or the yet to be identified subpopulations to gain a comprehensive understanding of the temperature-associated functions in *Daqu*.

Discussion

In this study, metagenomic sequencing, combined with the plate count method as well as physiological and biochemical indexes, were used to study temperature influenced the phage community in *Daqu* as well as its microbiome. The aim was to characterize and identify the similarities and differences in the microbial community structure, functions as well as physiological and biochemical indexes of *Daqu* at two temperatures, in order to reveal the peak temperature on the quality of *Daqu*. In addition, the key microorganisms during the brewing of *Daqu* and their respective functions at the two temperatures of *Daqu* were also revealed.

Several studies have proved that the regional climate, the water used in production and the microorganisms in the air jointly influence the microbial composition of *Daqu* (Hao et al. 2021). In particular, the peak and valley temperatures are one of the most important parameters that affect the growth and death of microorganisms, thereby determining the community structure and diversity of microorganisms in *Daqus* (Yan et al. 2013; Li et al. 2015b). Metagenomic sequencing has been increasingly used for profiling the microbiota of food products, and the development of this technique alongside bioinformatics enables us to better understand microbial composition and its metagenomic potential. Fine taxonomic resolution of food products microbiome also facilitates the identification and isolation of novel microbial resources (Leech et al. 2020).

Through metagenome sequencing, the microorganisms in two temperatures of *Daqu* were studied to determine their community structure and function. Using the bioinformatics software Bracken, the microbial composition of HD and MD were successfully revealed at the species level. The results further showed that the dominant species varied greatly between *Daqu* collected at different temperatures, with significant differences noted in terms of the microbial diversity and richness. For example, HD was dominated by *Kroppenstedtia eburnea* and *Sphingomonas* sp FARSPH, contrasting it to *Lactobacillus crustorum* and *Weissella cibaria* in MD. The metagenomic shotgun sequencing results suggested that HD and MD were different in their microbiota composition rather than level of complexity of the associated *Daqu* microbial communities. Molds and yeasts are mainly responsible for the alcoholic fermentation and saccharification in *Baijiu* fermentation (Fan et al. 2018). In this study, it was found that the number of molds and yeasts in MD was higher than that of HD. In addition, *Bacillus* spp were the dominant species in *Daqu*, and they are usually able to produce the most abundant metabolites to form the aroma of fermented


a
Genus
Kroppenstedtia
Lactobacillus
Phyllobacterium
Saccharopolyspora
Sphingomonas
Thermoactinomyces
Virgibacillus
Weissella
Suspected new species
Sample name
HD1
HD2
HD3
HD4
HD5
MD1
MD2
MD3
MD4
MD5
Completeness
Min. (82.8), Max. (99.7)
Contamination
Min. (0.0), Max. (7.0)
Tree scale: 0.1
b
MD4_6
HD1_7
MD1_2
MD2_5
HD2_6
HD4_6
HD1_9
HD2_2
MD4_1
HD2_4
HD5_2
HD4_4
HD1_5
HD3_3
c
AAs
CBMs
CEs
GHs
GTs
PLs
Relative abundance (%)
***
HD
MD
d
HD4_4
HD2_4
MD4_6
MD1_2
HD5_2
MD2_5
HD1_5
HD3_3
MD4_1
HD1_9
HD2_2
terpene
T3PKS
bacteriocin
ectoine
LAP
transAT−PKS
PKS−like
lassopeptide
betalactone
siderophore
NRPS
NRPS−like
TfuA−related
lanthipeptide
indole
e
Number of hit
KANAMYCIN
MACROLIDE
RIFAMPIN
RIFAMYCIN
SPECTINOMYCIN
VANCOMYCIN1
HD1_9
HD2_4
HD3_3
HD4_4
MD4_1

◂**Fig. 3** Assembly and characterization of *Daqu* metagenome-assembly genome (MAGs). **a** The phylogeny of 44 high-quality MAGs. The four MAGs that could not be assigned to a specific species are indicated with a yellow star; **b** Polysaccharide-degrading carbohydrate-active enzymes (CAZymes) encoded in the 14 representative SGBs; **c** CAZymes for the two groups of *Daqu*; **d** Secondary metabolite clusters encoded in the 14 representative SGBs; **e** Different classes of antibiotic-resistance genes (ARGs) detected in the SGBs. HD stands for high temperature *Daqu*; MD for medium–high temperature *Daqu*

food (Yan et al. 2013). As such, they could contribute to the evolution of flavor and enzyme activities such as amylases and proteinases needed for the fermentation of cooked sorghum for alcoholic fermentation (Fan et al. 2007).

Infection of fermentative bacteria is one of the main factors responsible for the failure of food fermentation (Gobbetti et al. 2018). Phage predation and subsequent horizontal gene transfer are natural ways for the biological control of microbial growth (Pujato et al. 2019) and influential to the *Daqu* microbiota composition and product quality, and the proportion of phage community was also extensively characterized in this study. The results revealed a large proportion of phage sequences. A significant negative correlation was observed between Podoviridae with *Weissella* and *Phyllobacterium*, indicating that the *Weissella* and *Phyllobacterium* host were killed by Podoviridae. In contrast, positive correlations were seen between Podoviridae and *Thermoactinomyces*, suggesting

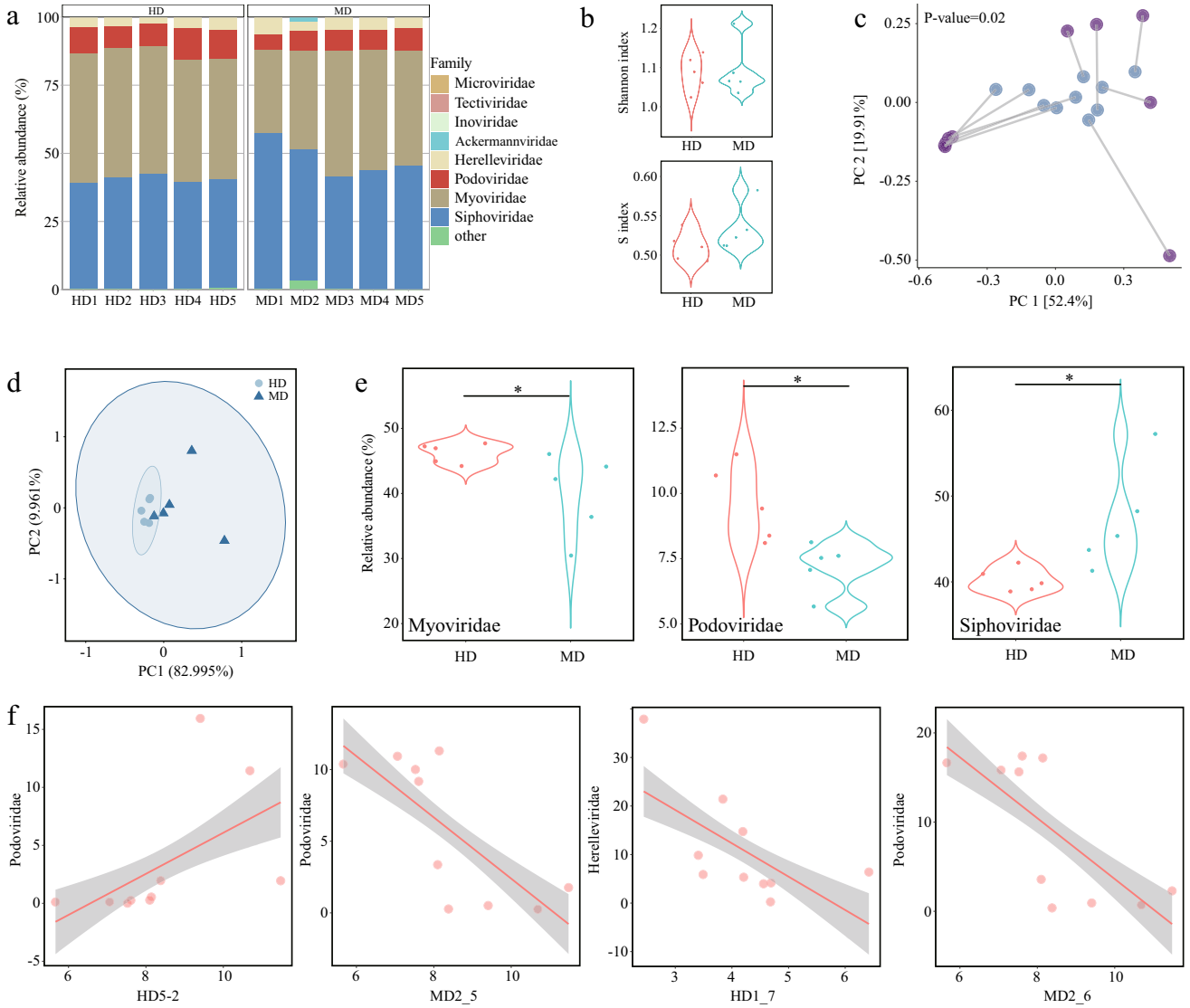


Fig. 4 Phage sequences in the *Daqu* microbiome. **a** Composition of major phage families in *Daqu*; **b** Boxplots showing values of Shannon index and S index between two sample groups; **c** Procrustes analysis of microbiota and phage of *Daqu*; **d** Principal coordinate analysis calculated based on Bray–Curtis distance (Phage microbiota); **e** Comparative box diagram of differential bacteriophages; **f** Correlations between SGBs and phage families detected in the cheese microbiome. HD stands for high temperature *Daqu*; MD for medium–high temperature *Daqu*

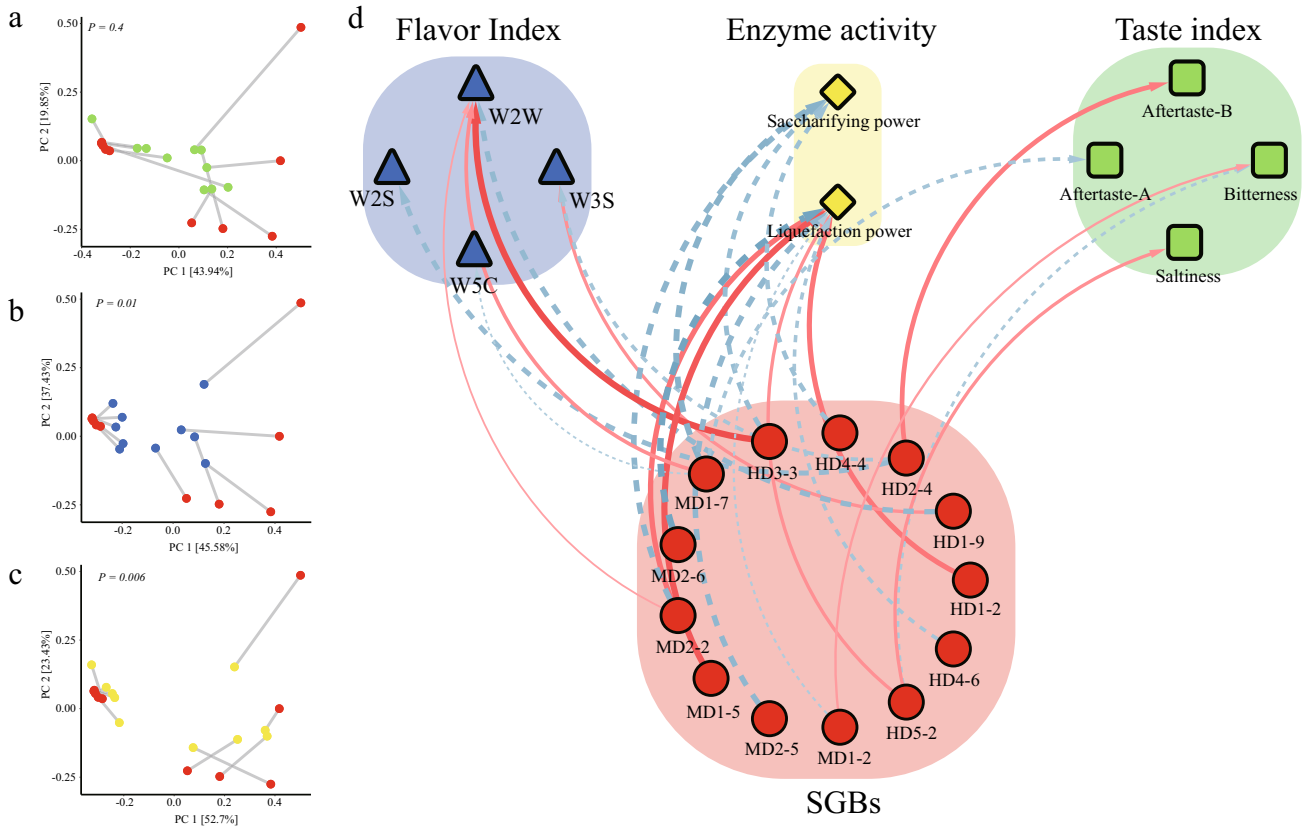


Fig. 5 Correlation between SGBs and specific functional components. **a** Procrustes analysis of microbiota and taste of *Daqu*; **b** Procrustes analysis of microbiota and flavor of *Daqu*; **c** Procrustes analysis of microbiota and enzyme activity of *Daqu*; **d** Correlation for SGBs with flavor, enzyme activity and taste (Spearman rank correlation test; $r > 0.6$, $P < 0.05$). The red circle, blue triangle, yellow diamond and green square represent SGBs, flavor, enzyme activity and taste, respectively. Red and blue lines represent positive and negative correlations, respectively. The darker the color, the stronger the correlation. Line thickness represents the magnitude of confidence level, with a thicker line representing a lower *p* value

more viable host-parasite interactions in these pairs and possibly a stable phage integration into the bacterial genomes. Besides, the metagenomic data suggested the likelihoof of horizontal transfer of ARGs from phages, hence further confirming the vital role of phages in shaping the microbial community structure of *Daqu*.

Different fermentation temperatures result in *Daqu* with different biochemical characteristics. At the same time, the quality of *Daqu* and its suitability for fermentation depend on its physiological and biochemical indexes. Generally, the *Daqu* moisture content should not be more than 13.0% to prevent excessive moisture resulting from the refermentation occurring during storage. The approximately 10.0% of moisture content measured in this study was in line with the requirements of *Daqu* production. Saccharifying ability is the ability of saccharifying microorganisms to convert starch into sugar and is one of the vital physiological and biochemical indexes of *Daqu* (Zheng et al. 2012). The results revealed that the saccharifying power of HD and MD was in the range of 350.0–650.0 U and 600.0–800.0 U, respectively. Related studies have confirmed that temperature has a great effect on enzyme activity and reaction rate (Xia et al. 2023). Therefore, the fermentation temperature may be the main reason for the above differences. Besides, the ability to liquefy starch into fermentable sugar directly is dependent on the quality of *Daqu* and is crucial during *Baijiu* production (Sha et al. 2015). The above experimental results showed that the liquefaction power of this HD was higher than that of MD.

Correlation analysis revealed the relationship between the microbial community structure and specific functional components of *Daqu*. The results revealed that the microbial community structure and its functions had significant effects on the specific functional components of *Daqu*. Some correlations could be explained by the functions of representative microbes of different temperatures of *Daqu*. For instance, *Saccharopolyspora* sp. with a higher relatively abundance in HD was significantly and positively correlated with aftertaste-B, but negatively correlated with W5C and

W3S. Similarly, Gan et al. showed that *Saccharopolyspora* sp was the dominant microorganism and was essential for producing flavor substances during Maotai liquor production (Gan et al. 2019). This could be due to the ability of the microorganism from this genus to produce erythromycin, with this process affecting the growth and metabolism of some microorganisms, while reduceing the flavor index utilization ability of the microbes. The results of this study fully provide an understanding of the microbial community, physiological and biochemical indexes as well as flavor substances of *Daqu*, which can provide a more comprehensive and scientific evaluation for the production suitability of *Daqu*.

Conclusion

This study focused on the differences and similarities in the microbial diversity, functions as well as physiological and biochemical indexes between *Daqu* from the two temperatures. Additionally, *Daqu* was comprehensively analyzed by means of strain-level metagenomic, quality analysis and physiological and biochemical property detection. The *Daqu* exhibited unique characteristics and a general set of indices that conformed to the general standard of the starter culture. It was further found that MD had a higher abundance of bacteria and fungi, compared with HD. The microbes with higher relative abundance in *Daqu* could not only produce specific flavor substances, but also had strong ability to degrade macromolecular substances in raw materials. A large number of phage sequences were even detected in the metagenomic dataset, suggesting that there were frequent and multiple interactions between the bacteria and phages. These results also provide evidence of stable genome integration and horizontal transfer of ARGs. The present study is significant for an improved understanding of *Daqu* and can serve as a guide the future exploration for the production and improvement of the final product quality.

Author contributions YW: Formal analysis, Writing—original draft, Writing—review and editing, Visualization. JG: Conceptualization, Resources. QH: Formal analysis, Software, Data curation. HZ: Project administration. CS: Supervision. ZG: Conceptualization, Funding acquisition.

Funding This work was financially supported by the Hubei Provincial Natural Science Foundation Joint Fund for Innovation and Development Project (2023AFD049) and Hubei University of Arts and Science Cultivation Fund for Teachers' Scientific Research Ability: Technological Innovation Team (2020kypytd009).

Data availability The sequence dataset has been deposited in the NCBI Sequence Read Archive (SRA) database (Accession number: PRJNA1010526, https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1010526).

Declarations

Competing interests The authors declare that they have no conflicts of interest.

References

Anyaso-Samuel S, Sachdeva A, Guha S, Datta S (2021) Bioinformatics pre-processing of microbiome data with an application to metagenomic forensics. In: Datta S, Guha S (eds) Statistical analysis of microbiome data. Springer, Cham, pp 45–78

Blin K, Kim HU, Medema MH, Weber T (2019) Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. Brief Bioinform 20:1103–1113

Chen L, Li Y, Jin L, He L, Ao X, Liu S, Zou L (2020) Analyzing bacterial community in pit mud of Yibin *Baijiu* in China using high throughput sequencing. PeerJ 8:9122

de Man TJ, Limbago BM (2016) SSTAR, a stand-alone easy-to-use antimicrobial resistance gene predictor. Msphere 1:50–65

Dong W, Yang Q, Liao Y, Liu Y, Hu Y, Peng N, Zhao S (2020) Characterisation and comparison of the microflora of traditional and pure culture Xiaoqu during the *baijiu* liquor brewing process. J I Brewing 126:213–220

Eddy SR (2011) Accelerated profile HMM searches. Plos Comput Biol 7:e1002195

Fan G, Sun B, Fu Z, Xia Y, Huang M, Xu C, Li X (2018) Analysis of physicochemical indices, volatile flavor components, and microbial community of a light-flavor *Daqu*. J Am Soc Brew Chem 76:209–218

Fan W, Xu Y, Zhang Y (2007) Characterization of pyrazines in some Chinese liquors and their approximate concentrations. J Agr Food Chem 55:9956–9962

Fang Z, Tan J, Wu S, Li M, Wang C, Liu Y, Zhu H (2020) PlasGUN: gene prediction in plasmid metagenomic short reads using deep learning. Bioinformatics 36:3239–3241

Gan SH, Yang F, Sahu SK, Luo RY, Liao SL, Wang HY, Liu H (2019) Deciphering the composition and functional profile of the microbial communities in Chinese Moutai liquor starters. Front Microbiol 10:1540

Gobbetti M, Di Cagno R, Calasso M, Neviani E, Fox PF, De Angelis M (2018) Drivers that establish and assembly the lactic acid bacteria biota in cheeses. Trends Food Sci Tech 78:244–254

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075

Hao F, Tan Y, Lv X, Chen L, Yang F, Wang H, Xu Y (2021) Microbial community succession and its environment driving factors during initial fermentation of Maotai-Flavor *Baijiu*. Front Microbiol 12:960

Hong J, Tian W, Zhao D (2020) Research progress of trace components in sesame-aroma type of *baijiu*. Food Res Int 137:109695

Hou Q, Wang Y, Cai W, Ni H, Zhao H, Zhang Z, Guo Z (2022) Metagenomic and physicochemical analyses reveal microbial community and functional differences between three types of low-temperature *Daqu*. Food Res Int 156:111167

Jin G, Zhu Y, Xu Y (2017) Mystery behind Chinese liquor fermentation. Trends Food Sci Technol 63:18–28

Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient

genome reconstruction from metagenome assemblies. PeerJ 7:7359
Kapli P, Yang Z, Telford MJ (2020) Phylogenetic tree building in the genomic age. Nat Rev Genet 21:428–444
Langdon WB (2015) Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. Biodata Min 8:1–7
Leech J, Cabrera-Rubio R, Walsh AM, Macori G, Walsh CJ, Barton W, Cotter PD (2020) Fermented-food metagenomics reveals substrate-associated differences in taxonomy and health-associated and antibiotic resistance determinants. Msystems 5:522–542
Li D, Liu CM, Luo R, Sadakane K, Lam TW (2015a) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
Li Z, Chen L, Bai Z, Wang D, Gao L, Hui B (2015b) Cultivable bacterial diversity and amylase production in two typical light-flavor *Daqus* of Chinese spirits. Front Life Sci 8:264–270
Li Z, Fan Y, Huang X, Han B (2022) Microbial diversity and metabolites dynamic of light-flavor *Baijiu* with stacking process. Fermentation 8:67
Liu F, Miao Y, Liu Y, Hou T (2020) RNN-VirSeeker: a deep learning method for identification of short viral sequences from metagenomes. IEEE ACM Trans Comput Biol 19:1840–1849
Maguire F, Jia B, Gray KL, Lau WYV, Beiko RG, Brinkman FS (2020) Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic islands. Microb Genomics 6:1–9
Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, Rasmussen S (2021) Improved metagenome binning and assembly using deep variational autoencoders. Nat Biotechnol 39:555–560
Nousias O, Montesanto F (2021) Metagenomic profiling of host-associated bacteria from 8 datasets of the red alga *Porphyra purpurea* with MetaPhlAn3. Mar Genom 59:100866
Olm MR, Brown CT, Brooks B, Banfield JF (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868
Pujato SA, Quiberoni A, Mercanti DJ (2019) Bacteriophages on dairy foods. J Appl Microbiol 126:14–30
Sha JX, Zheng DM, Liu JZ, Ning ML (2015) Detection the *Daqu* liquefying power by spectrophotometry. Liquor Mak 42:96–99
Sieber CM, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF (2018) Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3:836–843
Wang MY, Yang JG, Zhao QS, Zhang KZ, Su C (2019) Research progress on flavor compounds and microorganisms of Maotai flavor *baijiu*. J Food Sci 84:6–18
Wang Y, Xiang F, Zhang Z, Hou Q, Guo Z (2020) High-throughput sequencing-based analysis of fungal diversity and taste quality evaluation of Douchi, a traditional fermented food. Food Sci Nutr 8:6612–6620
Wang Y, Xiang F, Zhang Z, Hou Q, Guo Z (2021) Characterization of bacterial community and flavor differences of different types of Douchi. Food Sci Nutr 9:3460–3469
Xia Y, Luo H, Wu Z, Zhang W (2023) Microbial diversity in jiuqu and its fermentation features: saccharification, alcohol fermentation and flavors generation. Appl Microbiol Biot 107:25–41
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. J I Brewing 123:5–12
Yan Z, Zheng XW, Han BZ, Han JS, Nout MJ, Chen JY (2013) Monitoring the ecology of Bacillus during *Daqu* incubation, a fermentation starter, using culture-dependent and culture-independent methods. J Microbiol Biotechnol 23:614–622
Yang C, You L, Kwok LY, Jin H, Peng J, Zhao Z, Sun Z (2021) Strain-level multiomics analysis reveals significant variation in cheeses from different regions. LWT 151:112043
Yang C, Zhao F, Hou Q, Wang J, Li M, Sun Z (2020) PacBio sequencing reveals bacterial community diversity in cheeses collected from different regions. J Dairy Sci 103:1238–1249
Zheng XW, Yan Z, Han BZ, Zwietering MH, Samson RA, Boekhout T, Nout MR (2012) Complex microbiota of a Chinese "Fen" liquor fermentation starter (Fen-*Daqu*), revealed by culture-dependent and culture-independent methods. Food Microbiol 31:293–300
Zheng XW, Tabrizi MR, Nout MR, Han BZ (2011) *Daqu*-a traditional Chinese liquor fermentation starter. J I Brewing 117:82–90