




Analysis of microbial and metabolic diversity in Jiangshui from Northwest China

Haiyan ZHANG^{1,2} , Shuya XIANG^{1,2}, Ru ZHAI^{1,2}, Xuyang LI^{1,2}, Mingzhen HU^{1,2}, Tong WANG^{1,2},
Huiling ZHANG^{1,2}, Lin PAN^{1,2*}

Abstract

Jiangshui is a traditional fermented food in Northwest China with high nutritional value and special pharmacological effects. To study the microbial community structure and differences in the main metabolic pathways and metabolites in naturally fermented Jiangshui, the microbial diversity and nontargeted metabolomics analyses were performed on 18 samples of Jiangshui from three different production areas, namely, Tianshui, Gansu, Guyuan, Ningxia, and Ankang, Shaanxi. First, the microbial diversity was analyzed using 16S rRNA high-throughput sequencing. *Lactobacillus* was the dominant bacterial genus in the three regions, and the relative abundance of the samples in each region ranged from 72.69% to 99.95%. The dominant fungal genus was *Dipodascus*, with a relative abundance > 40%. Second, untargeted metabolomics was used to analyze the differences among metabolites in Jiangshui samples. Significant differences were found in 761 different metabolites detected in Jiangshui from the three different regions. Based on ESI⁺, 78 significantly different metabolites were screened using variable importance projection (VIP) value > 1, fold change (FC) < 1, and P-value < 0.05. Finally, based on joint analysis, the microbiome and metabolite groups of Jiangshui samples from the three different regions were evenly and closely clustered, and the microorganisms and metabolites were highly correlated. The results provide theoretical basis and reference for Jiangshui in other regions.

Keywords: Jiangshui; diversity microorganisms; metabolomics.

Practical Application: Jiangshui is a traditional fermented food in northwest China, which has high nutritional value and good flavor. The lactic acid bacteria in Jiangshui can degrade nitrite, and has special pharmacological effects, which can promote digestion, regulate viscera and urinate, and reduce cholesterol. It has become an indispensable part of the traditional fermented food diet in China. In this study, 16S rRNA high-throughput sequencing technology and untargeted metabolomics technology were used to determine the differences of microbial communities in different regions.

1 Introduction

Fermented vegetables have become an indispensable part of the Chinese traditional fermented food diet. Due to differences in regions and customs, fermented vegetables vary considerably and attract consumers due to their uniqueness. Jiangshui is a distinctive traditional fermented vegetable widely distributed in Northwest China (Hou et al., 2013; Lü et al., 2014). Jiangshui can improve taste, promote digestion, and prolong storage time. The lactic acid bacteria in Jiangshui can also degrade nitrite. The various raw ingredients are inexpensive, simple to process, and unique in flavor. Jiangshui tastes sour and mellow, is rich in nutrients, vitamin C₂, microbial B₂, organic acids (e.g., lactic acid, citric acid, malic acid), carotenoids, flavonoids, and sulfur amino acids, which are conducive to digestion. Jiangshui regulates qi, relieves thirst, boredom and lethargy, regulates viscera and urination, promotes digestion, and lowers cholesterol. Jiangshui can be directly used as a summer drink and as an ingredient for cooking and noodle soup (Cao et al., 2017; Li et al., 2022). However, the microbial community structure of Jiangshui is complex (Hou et al., 2013). Zhang et al. (2018) found that bacterial diversity changed with different Jiangshui.

Although microbial diversity has been reported in various types of fermented vegetables, such as kimchi in Korea, sauerkraut in Northeast China, and camuoi (fermented eggplant) in Vietnam (Nguyen et al., 2013), research investigating microbial populations involved in the production of Jiangshui is limited. Due to the different raw ingredients, processes, production conditions, and geographical differences in the production of Jiangshui, the microbial community of Jiangshui may also change accordingly (Zhang et al., 2018).

In the present study, the Jiangshui from Guyuan, Ningxia, Tianshui, Gansu, and Ankang, Shaanxi was used to analyze its microbial diversity using 16S rRNA high-throughput sequencing technology, explore differences in microbial classification, composition and abundance, and use untargeted metabolomics technology to determine the relationship between the related metabolic differential substances and metabolic pathways in the microorganisms. This study complements previous research and provides a reference for other regions and future research on the composition and content of Jiangshui metabolites.

Received 23 Oct., 2022

Accepted 09 Dec., 2022

¹School of Food and Wine, Ningxia University, Yinchuan, China

²Ningxia Key Laboratory for Food Microbial-Applications Technology and Safety Control, Ningxia University, Yinchuan, China

*Corresponding author: panlin@nxu.edu.cn

2 Materials and methods

The raw materials used in this experiment included six samples of fresh commercial Jiangshui (designated NX, SX, and GS, respectively) from Guyuan, Ningxia, Ankang, Shaanxi, and Tianshui, Gansu. The clear, transparent, and foam-free Jiangshui was placed into a sterile centrifuge tube and stored at low temperature before transporting to the laboratory. CTAB genomic DNA extraction kit, HPLC grade methanol, and acetonitrile were purchased from Thermo Fisher Scientific, formic acid from CNW, 2-propanol from Merck, 2-chloro-1-phenylalanine from Adamas-beta, AxyPrepDNA gel recovery kit from Axygen Co., Ltd., PCR amplification reagents from Bao Bioengineering (Dalian) Co., Ltd., and methanol from Tianjin Damao chemical reagent factory. In addition, PCR instrument (ABI GeneAmp®9700), QuantiFluor™ St blue fluorescence quantitative system (Promega), UHPLC system (Vanquish Horizon system), and Q-Exactive HF-X mass spectrometer were used for analyses.

2.1 16S rRNA high-throughput sequencing

DNA extraction of environmental samples and PCR amplification

The genomic DNA was extracted and detected using 1% agarose gel electrophoresis. The extracted Jiangshui DNA was diluted to 1 ng/μL as a template for PCR amplification. Bacteria amplified the V3-V4 region of 16S rDNA as the target DNA sequence, and fungi amplified the ITS1 region of the endogenous 18S rDNA-5.8S rDNA transcription interval as the target DNA sequence. Each sample had three replicates. After mixing the PCR products from the same sample, 2% agarose gel electrophoresis was performed to detect the purity and concentration of the PCR products. The AxyPrepDNA gel recovery kit was used to cut the gel to recover the PCR products.

Quantification and homogenization of PCR products

Based on the preliminary quantitative electrophoresis results, QuantiFluor™ St blue fluorescence quantitative system (Promega) was used for the detection and quantification of PCR products. Based on the sequencing volume requirements of each sample, the corresponding proportion was mixed and sent to Shanghai Majorbio Bio-pharm Technology Co., Ltd.

Miseq sequencing

Sequencing analysis was performed using the Miseq high-throughput sequencing platform. The original sequence data obtained were uploaded to the sequence reading Archive (SRA) of NCBI (accession number PRJNA844900).

2.2 Ultra-high Performance Liquid Chromatography (UHPLC) tandem Fourier-transform Mass Spectrometry (MS)

Sample handling

A total of 200 μL sample was placed into a 1.5 mL centrifuge tube and 800 μL extraction solution (methanol:acetonitrile = 1:1 (V:V))

containing 0.02 mg/mL internal standard (1-2-chlorophenylalanine) was added. The sample was vortexed for 30 s followed by low-temperature ultrasonic extraction for 30 min (5 °C, 40 kHz). The sample was left at -20 °C for 30 min then centrifuged at 13,000 g, 4 °C for 15 min. The supernatant was removed, dried with nitrogen, and 100 μL complex solution (acetonitrile:water = 1:1) was added for redissolution. The sample was vortexed for 30 s followed by low-temperature ultrasonic extraction for 5 min (5 °C, 40 kHz). Next, the sample was centrifuged for 10 min under the same centrifugation conditions and the supernatant transferred to the injection vial with internal cannula for analysis. In addition, 20 mL supernatant from each sample was mixed and used as the quality control (QC) sample. The QC sample was prepared by mixing the extracts of all samples in the same volume. The volume of each QC sample was the same as the sample volume. A QC sample was inserted every 5-15 analyses to investigate the stability of the entire detection process.

LC conditions

ACQUITY UPLC HSS T3 column (2.1 mm × 100 mm, 1.8 μm; Waters Corporation, Milford, USA) and UHPLC system combined with Q precision mass spectrometer were used for analysis. The mobile phase A was 95% water + 5% acetonitrile (containing 0.1% formic acid), mobile phase B was 5% acetonitrile + 47.5% isopropanol + 5% water (containing 0.1% formic acid), and the injection volume was 2 μL. The column temperature was 40 °C (Wang et al., 2022; Wang et al., 2019). The following elution gradient was used: 0-3.5 min, 0% B; 3.5-5 min, 24.5-65% B; 5-7.4 min, 65-100% B; 5-7.4 min, 65-100% B; 7.4-7.6 min, 100-51.5% B; 7.6-10 min, 0% B. The primary and secondary mass spectrometry data were collected using the UHPLC-Q-Exactive HF-X mass spectrometer of Thermo Fisher Scientific.

MS conditions

The samples were ionized by spray and the MS signals collected in the positive and negative ion scanning modes. The conditions of the ESI⁺ source were as follows: sheath gas flow, 50 Arb; auxiliary gas flow, 13 Arb; capillary temperature, 325 °C; full scan resolution, 60,000; MS2 resolution, 7,500; collision energy, 20/40/60 eV; ionization voltage, 3.5 kV (positive) and -3.5 kV (negative).

2.3 Bioinformatics analysis

16S rRNA high-throughput sequencing

Paired-end (PE) reads obtained from Miseq sequencing were first spliced based on the overlap relationship, and simultaneously, the sequence quality was controlled and filtered. After the samples were identified, OTU clustering and taxonomic analysis were performed. Then, diversity index analysis and sequencing depth detection were performed based on OTU and taxonomic information. Subsequently, statistical analysis of community structure was performed at various taxonomic levels and the community composition of multiple samples analyzed.

Untargeted metabolism

The raw data were imported into the metabolomics processing software ProgenesisQI (Waters Corporation) for baseline filtering, peak identification, integration, retention time correction, and peak alignment. A data matrix containing information such as retention time, mass-to-charge ratio, and peak intensity was obtained. Then the software was used to identify the characteristic peak search library, match the MS and MS/MS information with the metabolic database, set the MS mass error to < 10 ppm, and identify the metabolites based on the secondary MS matching score. The main databases used were the mainstream public databases and self-built databases (Li et al., 2021). Subsequently, the metabolites were subjected to multivariate statistical analyses, including principal component analysis (PCA), orthogonal partial least squares discriminant analysis (OPLS-DA), cluster analysis, and heat map analysis. Variable importance projection (VIP), fold change (FC), and P-values were used to screen the substances with significant differences between samples from different regions, and KEGG metabolic pathway enrichment analysis was performed on the metabolites with significant differences.

3 Results

3.1 16S rRNA high-throughput sequencing

Sample OTU analysis results

Based on high-throughput sequencing analysis of Jiangshui samples collected from the three different regions, the effective sequences of bacterial communities in Jiangshui samples from Tianshui, Gansu, Guyuan, Ningxia, and Ankang, Shaanxi were 75,908, 62,704, and 83,439, respectively. OTU cluster analysis

was performed on the obtained sequences and a total of 55 OTUs were obtained. As shown in Figure 1A, all OTUs were divided into 3 phyla, 4 classes, 5 orders, 5 families, and 6 genera at different classification levels. The effective sequences of fungal communities in Tianshui, Gansu, Guyuan, Ningxia, and Ankang, Shaanxi were 112,081, 76,443, and 85,099, respectively. As shown in Figure 1B, 347 OTUs were obtained by clustering all sequences, which were divided into 4 phyla, 14 classes, 32 orders, 68 families, and 100 genera at different classification levels.

Analysis results of microbial community structure

The proportion of each species in the bacterial community of Jiangshui samples from the three regions at the genus classification level is shown in Figure 2A. At the genus classification level, the bacterial community mainly belonged to *Lactobacillus* and *Acetobacter*. *Lactobacillus* was determined the core bacterial group of the bacterial community in the Jiangshui fermentation process and plays an important role in the formation of its sour taste (Chen et al., 2016a). As shown in Figure 2B, the genus *Dipodascus* was the dominant population with the highest abundance in the Jiangshui samples from various regions; however, the abundance of different genera in each sample significantly varied. In general, the fungal communities in the Jiangshui samples from the three regions showed significant differences in the composition and abundance of dominant flora.

Rank curve and cluster curve analysis results

Rarefaction and Rank abundance clustering were used to analyze species abundance and evenness. The changes in bacterial communities in the Jiangshui samples from the different regions

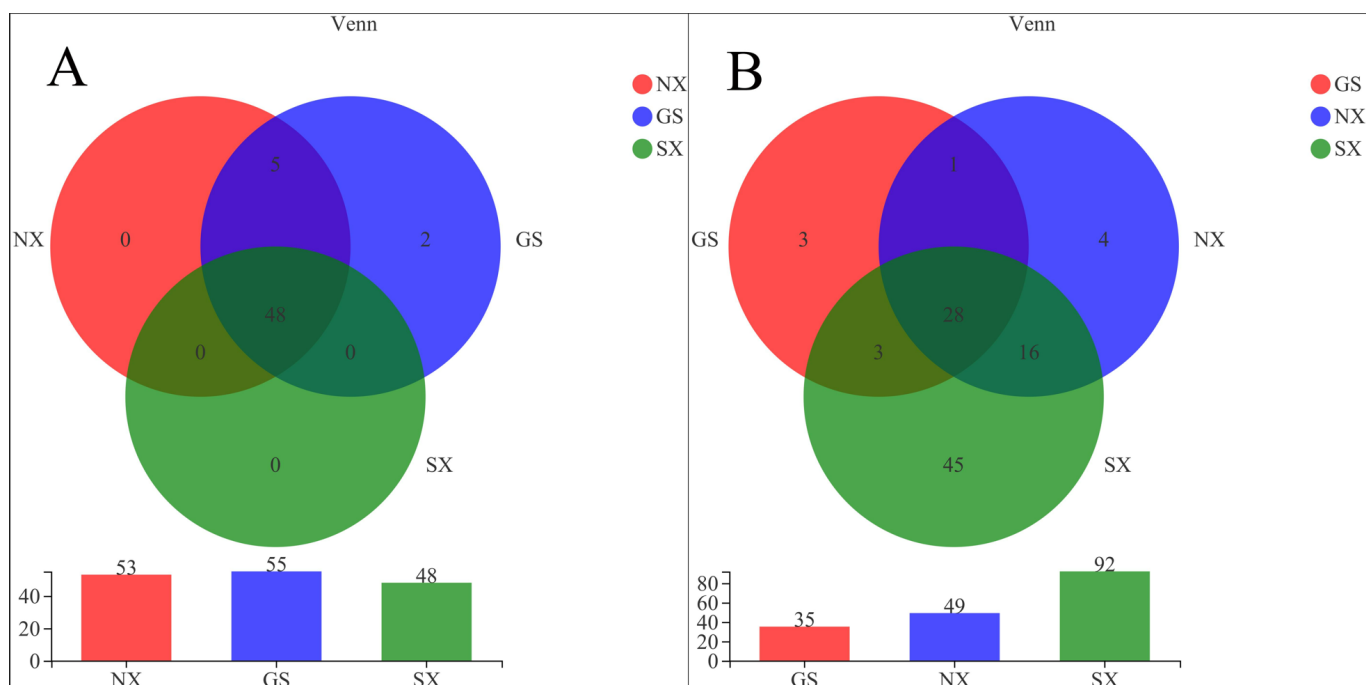


Figure 1. Bacterial 16S rRNA (A) in Jiangshui from three different regions at OTU (operational taxonomic units) level Venn diagram of fungal ITS (B) and column diagram of the total number of species in each group.

are shown in Figure 3A. The grade curve shows the bacterial species richness in most of the Jiangshui samples was from 10-40 at the OTU level. The Jiangshui samples from Tianshui,

Gansu had the widest distribution range and the highest species diversity, and the dominant bacterial groups in the Jiangshui samples from Ankang, Shaanxi had a high proportion and

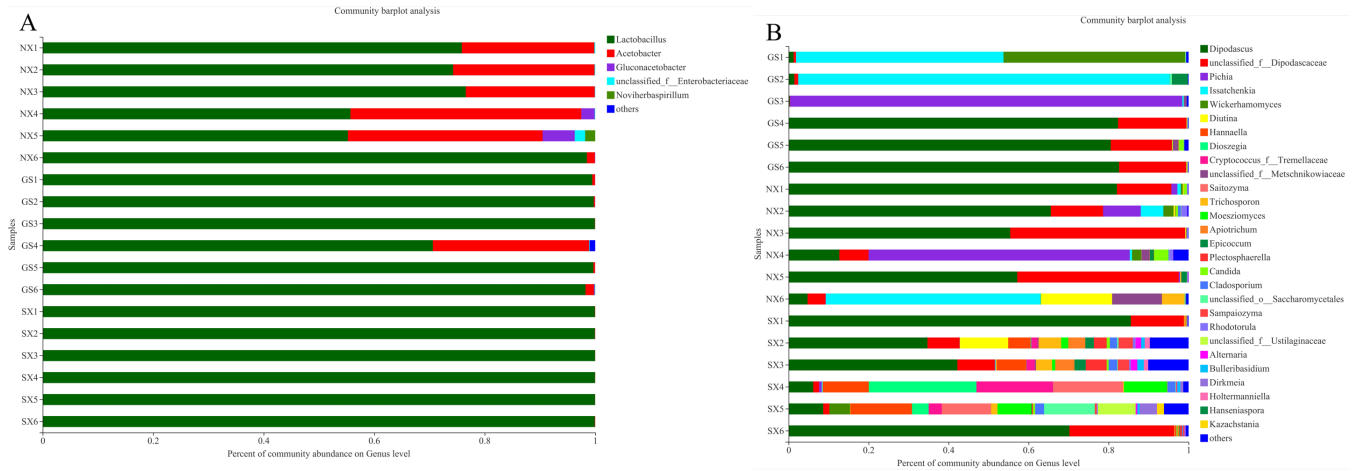


Figure 2. Column chart of bacterial community (A) and fungal community (B) structure in Jiangshui from three different areas. The ordinate is the sample name, the abscissa the proportion of the species in the sample, the columns of different colors represent different species, and the length of the columns represents the proportion of the species.

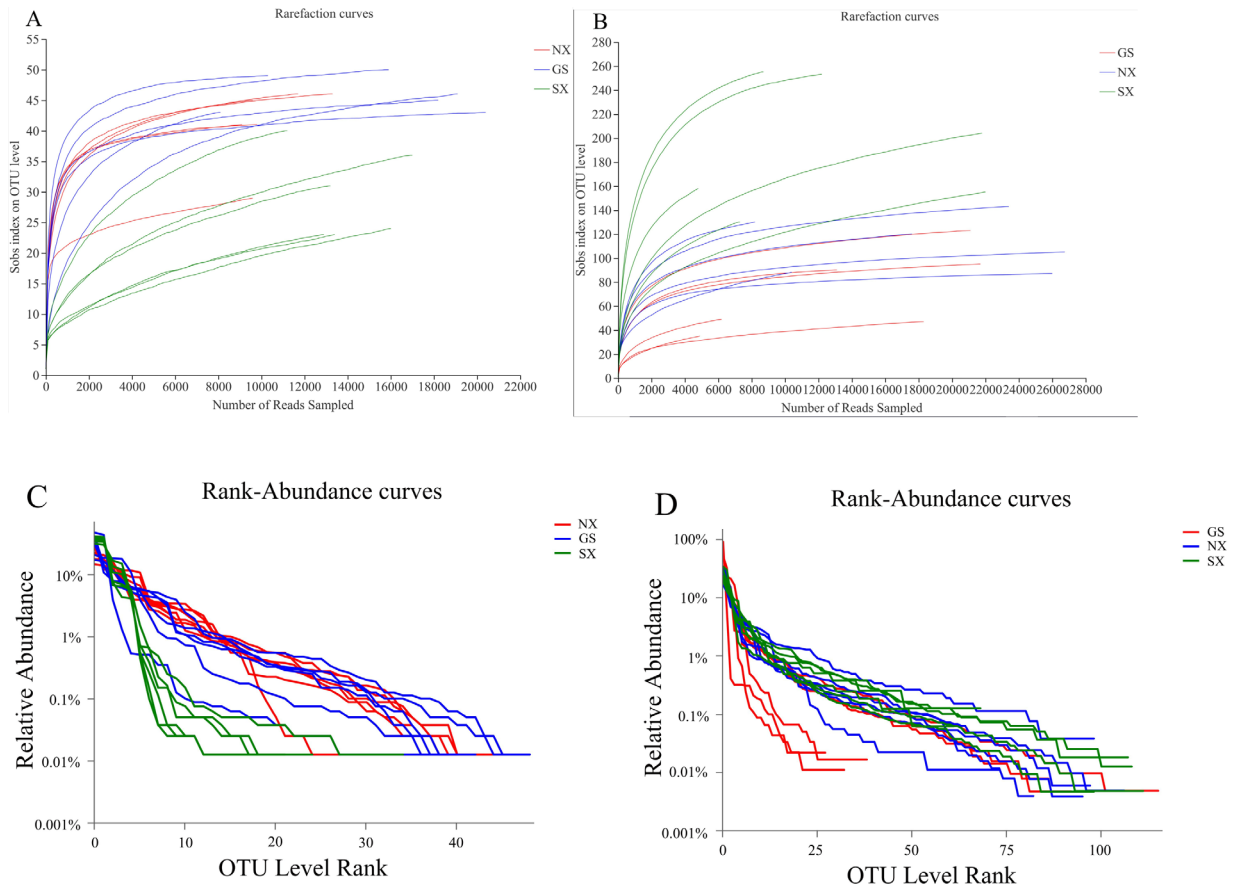


Figure 3. Rarefaction curves and Rank-Abundance curves of microbial communities in Jiangshui from different regions. The abscissa represents the ranking level of the number of species under the OTU taxonomic level, the ordinate represents the relative percentage content of the number of species under the taxonomic level, and the abscissa position of the extension end point of the sample curve is the number of species of the sample. (A) Bacterial community Rarefaction curves; (B) Bacterial community Rank-Abundance curves; (C) Fungal community Rarefaction curves; (D) Fungal community Rank-Abundance curves.

low diversity. As shown in Figure 3B, the species richness of the fungal community was mostly from 20-100. Most of the Jiangshui samples from Ankang, Shaanxi, Guyuan Ningxia, and a small portion of the Jiangshui samples from Tianshui, Gansu showed an overall steady downward trend, indicating the species diversity was high. The curve for most of the Jiangshui samples from Tianshui, Gansu sharply decreased, indicating the dominant fungal flora was prominent and diversity was the lowest. As shown in Figure 3C, the dilution curve for the bacterial community in the Jiangshui sample tended to be flat after the serial number was 16,000. That for the fungal community tended to be stable after the sequence number was 22,000. With an increase in sequencing depth, the curve tended to stabilize, indicating that sequencing included most of the sample information. Therefore, the dilution curve showed the sequencing data of microorganisms in Jiangshui was reasonable.

Microbial community α diversity analysis results

As shown in Table 1 and 2, the ACE and Chao1 indices show the microbial abundance of the bacterial community in Tianshui, Gansu was the largest and the fungal community in Ankang, Shaanxi was the largest. The Shannon and Simpson index showed the diversity of bacterial communities in Guyuan, Ningxia and Ankang, Shaanxi was higher, and the diversity of fungal communities in Ankang, Shaanxi and Tianshui, Gansu was higher. The sequencing depth index of Jiangshui samples from the three different regions was > 99%, indicating most species could be detected, the sample sequencing met the requirements, and sequencing results represent the actual condition of microorganisms in the samples.

Microbial community β diversity analysis results

As shown in Figure 4A, the explanatory value of the difference in sample composition was > 50% and the Jiangshui

samples in Ankang, Shaanxi, were closely clustered to each other, indicating the bacterial microbial community composition of the Jiangshui samples from Ankang, Shaanxi, was highly similar, and the Jiangshui samples from Tianshui, Gansu, and Guyuan, Ningxia, were similar, indicating the species composition between the two regions was relatively similar. In addition, the fungal microbial community composition of several samples from the three different regions was highly similar, and the fungal species composition of the samples was generally different as shown in Figure 4B. These results indicate bacterial microbial and fungal microbial communities were significantly different in Jiangshui from different areas.

Table 1. Detailed information for the samples used in the study.

Sample	Group	Source
Gansu	GS1	Tianshui
	GS2	Tianshui
	GS3	Tianshui
	GS4	Tianshui
	GS5	Tianshui
	GS6	Tianshui
Ningxia	NX1	Guyuan
	NX2	Guyuan
	NX3	Guyuan
	NX4	Guyuan
	NX5	Guyuan
	NX6	Guyuan
Shaanxi	SX1	Ankang
	SX2	Ankang
	SX3	Ankang
	SX4	Ankang
	SX5	Ankang
	SX6	Ankang

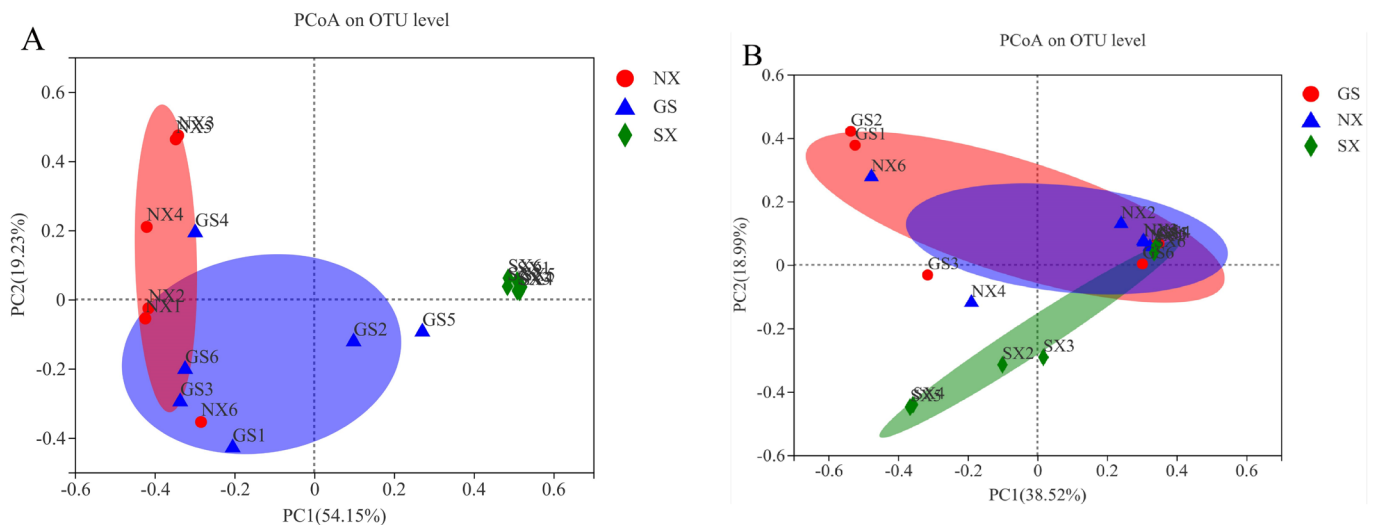


Figure 4. PCoA (principal coordinates analysis) map of microbial community diversity in Jiangshui from different regions. The x-axis and y-axis represent the two selected principal axes and the percentage represents the explanatory value of the principal axis to the difference in sample composition; the scale of x-axis and y-axis is a relative distance which has no practical significance. (A) Bacterial community; (B) Fungal community.

Table 2. Alpha diversity index of Jiangshui in different regions.

Sample/ Estimators	ACE		Chao1		Shannon		Simpson		Coverage (%)	
	16S	ITS	16S	ITS	16S	ITS	16S	ITS	16S	ITS
GS1	47.69	68.42	48	69	2.41	1.88	0.15	0.19	99.98	99.74
GS2	50.32	49.10	49.5	50.6	1.13	1.41	0.42	0.35	99.96	99.73
GS3	45.94	128.20	44.5	81	2.24	1.51	0.20	0.31	99.99	99.91
GS4	49.33	106.54	49	106.14	2.76	2.39	0.09	0.17	99.99	99.94
GS5	47.72	137.21	47.2	134.77	2.10	2.53	0.16	0.17	99.91	99.91
GS6	53.59	94.09	51.2	94	2.29	2.41	0.20	0.17	99.97	99.93
NX1	42.26	138.10	41.3	158.5	2.59	2.45	0.13	0.19	99.98	99.87
NX2	45.55	157.76	42.5	166.33	2.58	2.82	0.12	0.13	99.97	99.91
NX3	40.48	92.70	34	92.6	2.49	2.51	0.10	0.14	99.95	99.97
NX4	47.54	144.52	55	161.63	2.45	2.59	0.13	0.19	99.95	99.72
NX5	48.82	114.76	47.5	114.75	2.49	2.54	0.11	0.16	99.97	99.95
NX6	47.65	115.79	46.75	109.43	2.45	2.72	0.13	0.12	99.98	99.76
SX1	42.30	266.10	51.5	236.67	1.30	2.21	0.34	0.22	99.93	99.77
SX2	45.50	274.93	42.8	276.25	1.52	4.13	0.29	0.04	99.93	99.59
SX3	39.90	262.54	37	262.04	1.52	3.97	0.26	0.05	99.93	99.81
SX4	33.44	163.11	41	156.71	1.53	3.05	0.25	0.08	99.93	99.53
SX5	28.10	185.49	26	197.18	1.42	3.80	0.29	0.03	99.95	99.23
SX6	49.20	258.05	47	259.43	1.32	2.72	0.34	0.14	99.93	99.77

ACE: the index used to estimate the number of OTUs in a community is one of the common indices used to estimate the total number of species in ecology; ITS: internal transcribed spacer sequence refers to the endogenous transcriptional interval between the 5.8 s, 18 s and 28 s rDNA sequences on fungal rDNA.

3.2 Untargeted metabolomics

PCA results

To determine the overall distribution trend of Jiangshui samples from different regions, PCA was used to explore the correlation between samples. Under ESI⁺ and ESI⁻ conditions, the sum of the principal component contribution values of the PCA maps of the three different regions was > 50%, indicating the fitting of this model was reliable. As shown in Figure 5 and Figure S1, the similarity was high between Jiangshui samples from Guyuan, Ningxia, and Tianshui, Gansu, but significantly different from Ankang, Shaanxi, samples. In addition, as shown in Figure 5, the separation trend among the three regions was minimal with no abnormal points, indicating species composition structure from the three different regions was highly similar.

OPLS-DA and permutation test results

Based on ESI⁺ and ESI⁻, OPLS-DA analysis results of Jiangshui samples from the three different regions are shown in Figure 6 and Figure S2. The samples of Jiangshui groups from the three different regions were evenly distributed on both sides of the origin, with clear boundaries and obvious separation, and large differences were observed between groups. To verify the reliability of the OPLS-DA model, $R^2X = 0.448$, $R^2Y = 1$, $Q^2 = 0.737$, and Q^2 were all > 0.5 and close to 1, indicating the model had good stability and prediction ability. Similarly, $R^2X = 0.607$, $R^2Y = 0.966$, $Q^2 = 0.759$, and $R^2X = 0.613$, $R^2Y = 0.996$, $Q^2 = 0.808$, indicating that the data of the model were reliable and can be used for subsequent screening analysis of metabolic difference.

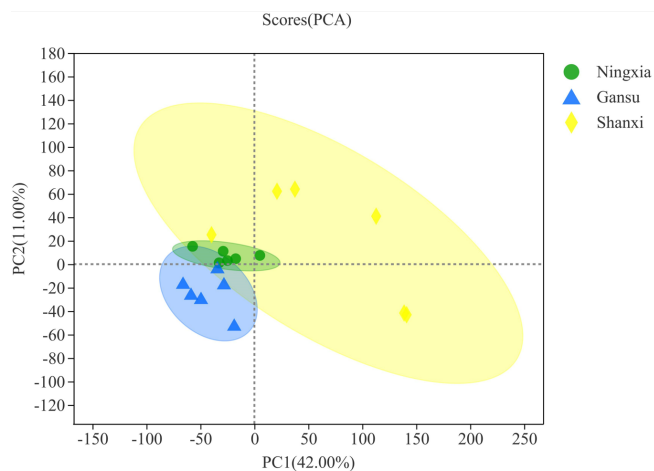


Figure 5. PCA scores of Jiangshui samples from different regions using ESI⁺. After dimension reduction analysis, relative coordinate points on the principal components p1 and p2 are found. The confidence ellipse indicates the “real” samples of this group were distributed in this area with 95% confidence.

Metabolite Venn analysis and differential metabolite cluster analysis results

As shown in Figure 7, there were 167 metabolites in the metabolic sets from Tianshui, Gansu, and Guyuan, Ningxia, 587 metabolites from Ankang, Shaanxi, and Tianshui, Gansu, 497 metabolites from Ankang, Shaanxi, and Guyuan, Ningxia, and 29 metabolites from the three regions. To determine the correlation and concentration change trend in different metabolites,

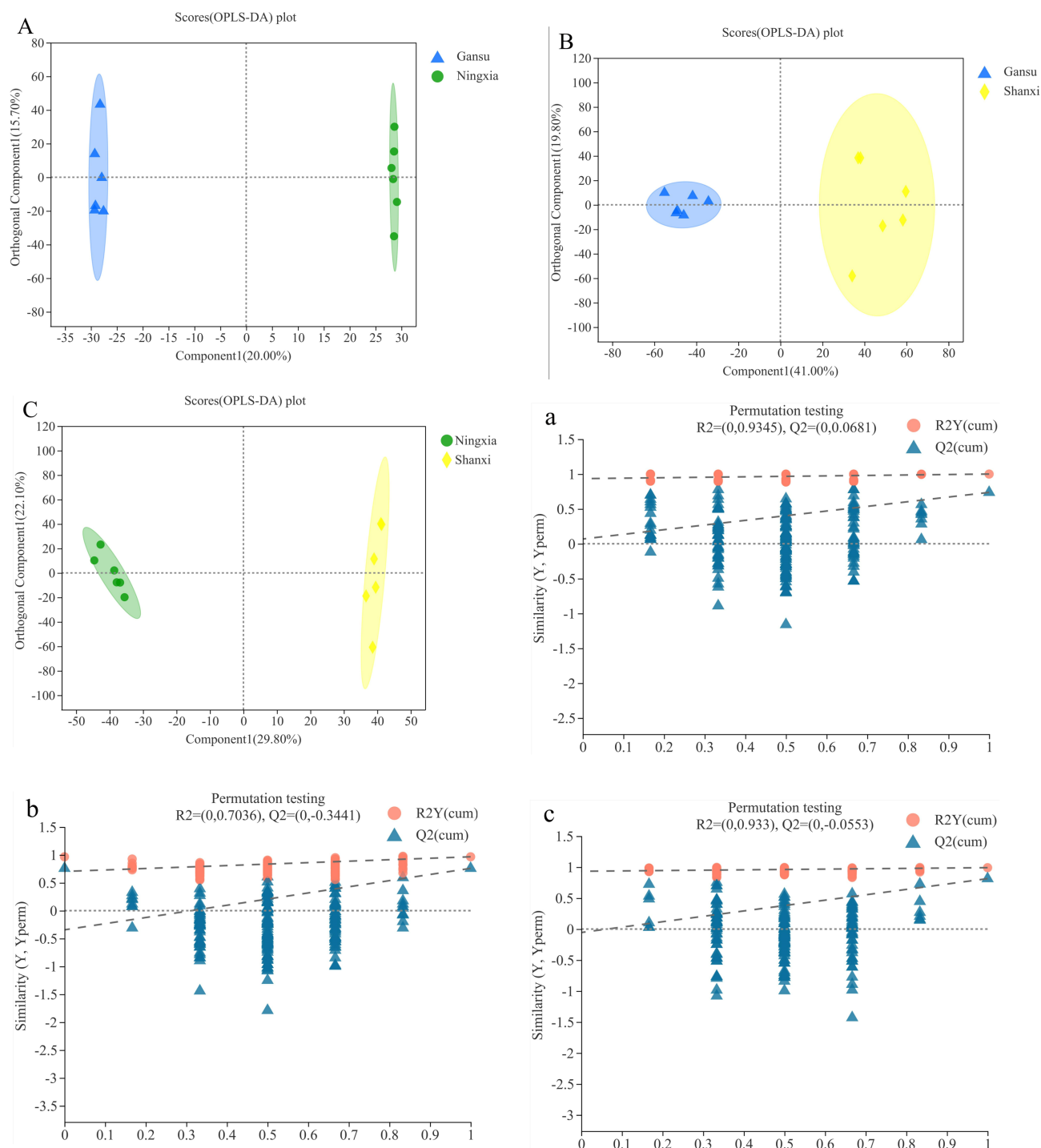


Figure 6. (A-C) are the OPLS-DA scores using ESI+, the OPLS-DA score map filters out the information irrelevant to the group through orthogonal rotation to better distinguish the differences between groups and improve the efficiency of the model. Comp1 is the first prediction principal component decomposition degree and orthogonal Comp1 is the first orthogonal component decomposition degree. (A-C) are the OPLS-DA model validation; the abscissa represents the displacement retention degree of the displacement test (the proportion consistent with the order of Y variables of the original model, and the point with the displacement retention degree of 1 is the R2 and Q2 values of the original model), the ordinate represents the values of R2 (red circle) and Q2 (blue triangle) displacement test, and the two dashed lines, respectively, represent the regression lines of R2 and Q2. A and a: GS vs. NX; B and b: GS vs. SX; C and c: NX vs. SX.

the metabolites with the highest 30 abundances were selected. As shown in Figure 8, the abundance of metabolites in Jiangshui

from Ankang, Shaanxi, was significantly greater compared to Guyuan, Ningxia, and Tianshui, Gansu. Based on hierarchical

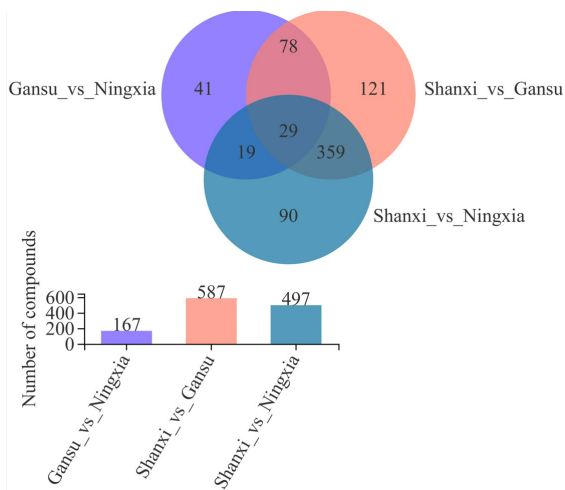


Figure 7. Venn diagram of each metabolic set of Jiangshui from the different regions. The overlapping area in the figure represents the number of metabolites common to multiple metabolic sets, the area without overlap represents the number of metabolites unique to the metabolic set, and the number represents the corresponding number of metabolites. The second histogram shows the number of metabolites contained in each metabolic set.

cluster analysis, the content of various metabolites in Jiangshui showed significant regional differences.

Analysis results of differential volcano map

Based on ESI⁺, as shown in Figure 9 and Figure S3, 453 differential metabolites were detected in all Jiangshui samples from the three different regions and 119 were differential metabolites in samples from Tianshui, Gansu, and Guyuan, Ningxia; 28 were upregulated and 91 were downregulated. The number of differential metabolites from Ankang, Shaanxi, and Guyuan, Ningxia, was 291; 250 were upregulated and 41 were downregulated. The number of differential metabolites from Ankang, Shaanxi, and Tianshui, Gansu, was 351; 303 were upregulated and 48 were downregulated. Similarly, based on ESI⁻, 284 differential metabolites were detected in Jiangshui samples from the three regions. The results showed the metabolites in Jiangshui samples from the three regions were significantly different and the identification results are shown in Table S1.

Screening and analysis of differential metabolites

Based on analysis results of the OPLS-DA model, metabolites with $P < 0.05$, $VIP > 1$, and $FC < 1$ based on t-test were used as

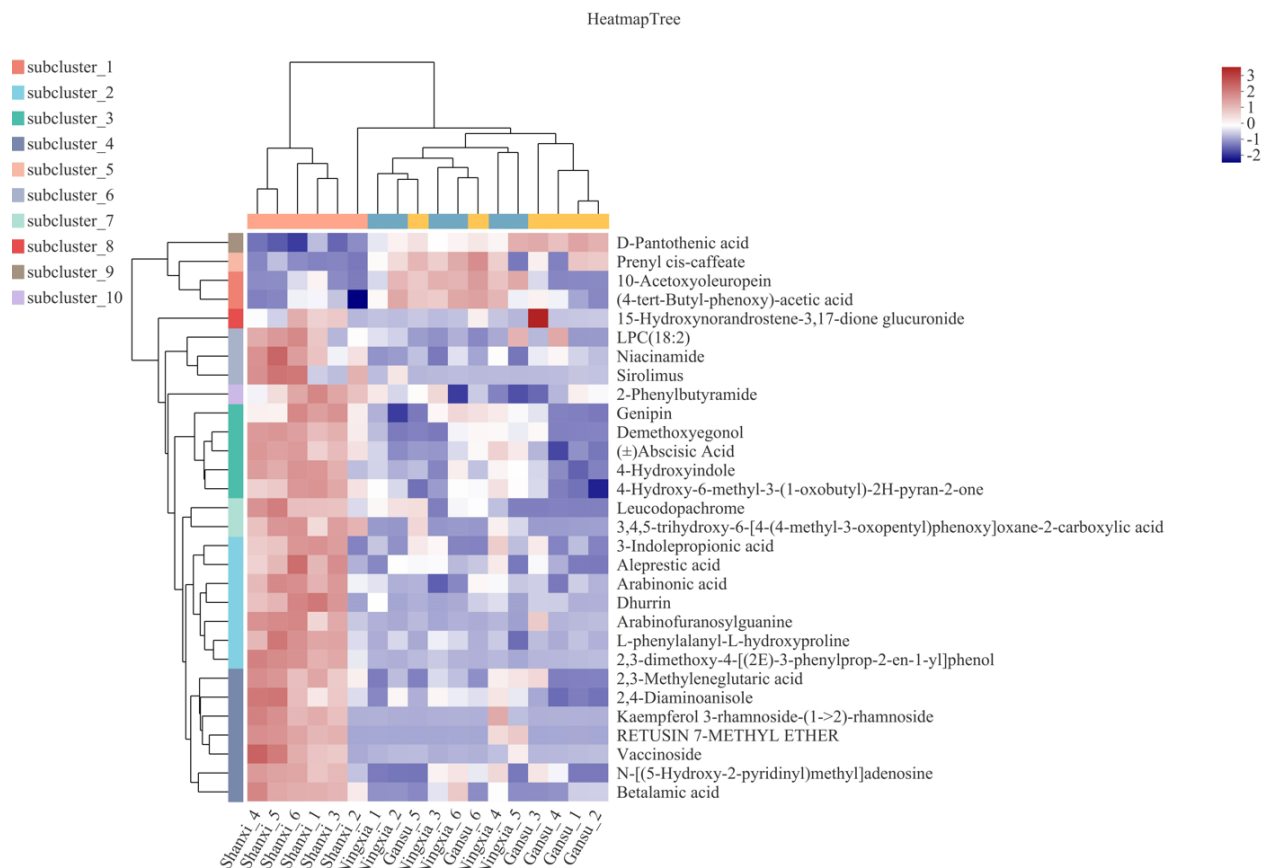


Figure 8. Cluster heatmap of metabolite cluster analysis among metabolic sets from the different regions. Each column in the figure represents a sample and each row represents a metabolite. The color represents the relative expression amount of metabolites in the group of samples. The left side is the tree diagram of metabolite clustering, the right side is the name of metabolites, the upper part is the tree view of sample clustering, and the lower part is the name of the sample.

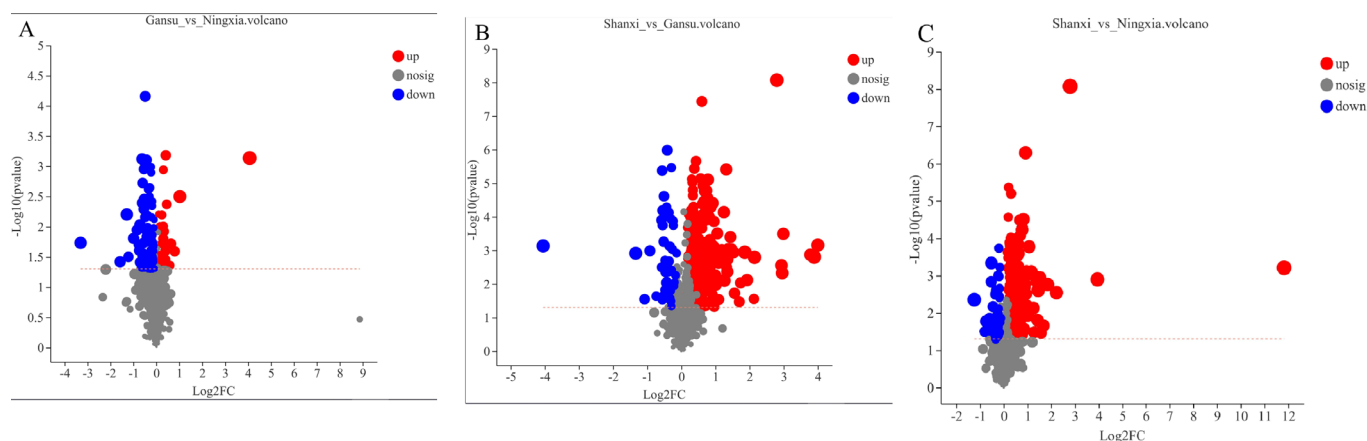


Figure 9. ESI⁺ volcano map of the difference between the Jiangshui groups from different regions. The abscissa is the FC value of the metabolite expression difference between the two groups, (i.e., log₂FC) and the ordinate is the statistical test value of the metabolite expression difference (i.e., -log₁₀ (P-value)). The values of the abscissa and the ordinate are logarithms. Each dot in the graph represents a specific metabolite and the size of the dot represents the VIP value. The points on the left are metabolites with differentially downregulated expression and the points on the right are metabolites with differentially upregulated expression.

screening criteria. Based on the retention time and mass-to-charge ratio data, the differential metabolites were searched using ESI⁺ and the identification results are shown in Table 3. A total of 180 differential metabolites were screened out from Jiangshui samples from the different regions. A total of 94 differential metabolites were screened using ESI⁻.

Differential metabolite pathway analysis

Pathway enrichment analysis was performed on the different metabolites based on the KEGG database. The differential metabolites were mainly distributed in 20 metabolic pathways. As shown in Figure 10, compared to Guyuan, Ningxia, Jiangshui from Tianshui, Gansu, had more pathways for protein digestion and absorption, aminoacyl-tRNA biosynthesis, central carbon metabolism in cancer, cyanoamino acid metabolism, biosynthesis of plant secondary metabolites, and cocaine addiction; the bubbles at protein digestion and absorption and aminoacyl-tRNA biosynthesis were the darkest and relatively large. A total of 7 detected differential metabolites were enriched in this metabolic pathway, including L-tyrosine, L-threonine, L-valine, histidine, phenylalanine, L-tryptophan, and L-glutamate.

3.3 Combined analysis of microbial diversity and metabolomics in Jiangshui

Results of multiple regression analysis

O2PLS is predictive bi-directional modeling tool. Figure 11A shows the close average cluster around the origin. Two samples (Tianshui, Gansu 4, and Guyuan, Ningxia 4) deviated from the average cluster. Figure 11B and Figure S4 shows the Jiangshui samples from the two regions were relatively discrete, and the data matrix of microbiome and metabolome were closely separated, indicating the internal correlation between the microbiome and metabolome in samples from Tianshui, Gansu, and Guyuan, Ningxia, was the strongest.

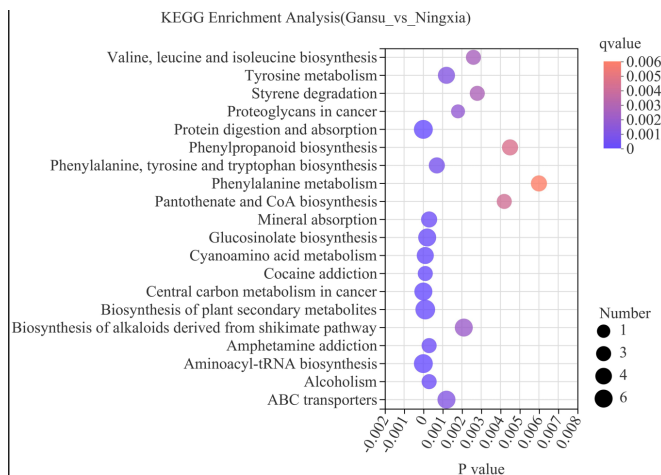


Figure 10. Bubble diagram of KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis for samples from GS and NX. The abscissa is the enrichment significance P-value and the ordinate is the KEGG pathway. The size of bubbles represents the amount of metabolites enriched in the metabolic set in this pathway.

Results of weighted gene coexpression network analysis

To further explore the relationship between microbiota and metabolome composition, the abundance of metabolites in Jiangshui samples from the three different regions was analyzed. As shown in Figure 12A and Figure S5, the correlation between each metabolite and microorganism was very high, and the positive correlation was higher than the negative correlation. A single metabolite/microorganism correlated with multiple microorganisms. Overall, there was a high correlation between the microbiome and metabolites.

4 Discussion

Jiangshui is a traditional fermented food widely consumed in Northwest China including Tianshui, Gansu, and Ankang,

Table 3. Difference of metabolites in Jiangshui samples from different regions under ESI⁺.

M/Z	Metabolite	Molecular Formula	VIP	Fold change	P-value
165.0758211	Dulcitol	C6H14O6	2.029510943	0.785977181	0.04619
182.0813976	L-Tyrosine	C9H11NO3	1.43072621	0.909182644	0.0347
247.1081591	N-Acetyl-DL-tryptophan	C13H14N2O3	3.477923069	0.409230769	0.006273
446.1641876	Neolinustatin	C17H29NO11	2.188140041	0.655991165	0.03146
181.0862142	Cinnamic acid	C9H8O2	1.044166887	0.93879056	0.007683
260.092448	4-Hydroxy-6-methyl-3-(1-oxobutyl)-2H-pyran-2-one	C10H12O4	1.512029004	0.842019984	0.02607
334.1406239	Hypoglycin B	C12H18N2O5	2.233698828	0.662004662	0.0272
191.1433524	4-(2,6,6-Trimethylcyclohexa-1,3-dienyl)but-3-en-2-one	C13H18O	2.039552122	0.754730713	0.01482
178.0865478	2'-Hydroxyacetophenone	C8H8O2	1.891102159	0.734065345	0.04353
130.0653821	Isoquinoline	C9H7N	2.71673368	0.741964127	0.0007863
287.0556033	Elenaic acid	C11H14O6	2.436665482	0.693766295	0.003888
302.1006546	Citrusinine I	C16H15NO5	2.238737254	0.72616408	0.0007893
176.0710376	3-Methoxytyrosine	C10H13NO4	1.234342591	0.912790698	0.02443
102.0553861	L-Threonine	C4H9NO3	1.294403772	0.921037998	0.02015
145.1700286	DL-Coniine	C8H17N	2.949032467	0.693620046	0.0008137
281.08029	Gamma-Glutamyl-S-methylcysteine sulfoxide	C9H16N2O6S	2.979431423	0.649488671	0.0007583
138.0550266	Trigonelline	C7H7NO2	1.492143073	0.910183428	0.02525
228.0980944	Asparaginyln-Hydroxyproline	C9H15N3O5	2.650068037	0.692497431	0.003499
325.1431368	7-Aminoflunitrazepam	C16H14FN3O	1.715185011	0.849871262	0.02426
168.065635	3alpha,4,5,7alpha-Tetrahydro-5-hydroxy-1H-isoindole-1,3(2H)-dione	C8H9NO3	1.21479967	0.923331952	0.02111
229.1187425	Prolylhydroxyproline	C10H16N2O4	1.584290885	0.864428389	0.02776
156.0769882	L-Histidine	C6H9N3O2	2.449054065	0.687629938	0.001114
203.139441	ALANYL-dl-LEUCINE	C9H18N2O3	1.898923585	0.730629371	0.03388
324.1086528	L-DOPA 3'-glucoside	C15H21NO9	1.552643218	0.881612091	0.003824
291.0392706	N5-(3,4-Dioxo-1,5-cyclohexadien-1-yl)-L-glutamine	C11H12N2O5	2.175234266	0.652560084	0.04625
427.7121853	Pimecrolimus	C43H68ClNO11	2.219307517	0.693689746	0.04594
294.1980295	Canthaxanthin	C40H52O2	1.85867212	0.72040521	0.04045
175.0392031	1-Oxo-1H-2-benzopyran-3-carboxaldehyde	C10H6O3	1.214054354	0.892444444	0.01691
229.155128	3-propan-2-yl-2,3,6,7,8,8a-hexahydropyrrolo[1,2-a]pyrazine-1,4-dione	C10H16N2O2	2.03055264	0.681867832	0.04289
158.0095975	Phosphocreatinine	C4H8N3O4P	2.87048892	0.502759659	0.01559
527.1281352	Egonol glucoside	C25H28O10	1.784131462	0.783887162	0.0482
470.106312	Theaflavic acid	C21H16O10	2.019039917	0.801032566	0.01736
148.0251054	Ibervirin	C5H9NS2	1.645077148	0.850324471	0.03456
207.0655953	CITROPTEN	C11H10O4	2.065397839	0.625273694	0.04162
172.0970128	(4S,6S)-3,4,5,6-Tetrahydro-4-hydroxy-6-methyl-2H-pyran-2-one	C6H10O3	1.511703978	0.834556033	0.04944
341.127105	Methionyl-Glutamine	C10H19N3O4S	2.894892226	0.334309991	0.03801
332.1111582	R-95913	C18H18FNO2S	2.119244342	0.740592474	0.004474
175.1485095	1,2,3,4,tetrahydro-1,5,7-trimethylnapthalene	C13H18	1.152057235	0.895678092	0.03137
178.0357795	DL-Sulforaphane	C6H11NOS2	1.951249234	0.748321313	0.03288
215.0311155	8-Chlorotheophylline	C7H7ClN4O2	1.958022247	0.734320557	0.0146
548.3091592	Cinnassiol D2 glucoside	C26H42O11	2.588100354	0.68540038	0.005115
162.055218	4-Hydroxy-2-quinolone	C9H7NO2	2.010428479	0.761569416	0.01143
209.0812365	5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone	C11H12O4	1.38752295	0.832493703	0.04378
508.1548089	L-2-Amino adipate adenylate	C16H23N6O10P	2.441876124	0.714913244	0.004083
215.1183237	7-Hydroxytycaine	C13H14N2O	2.025336858	0.682776061	0.04476
261.1061423	Hydroxypropyl-Glutamate	C10H16N2O6	3.400463514	0.101367684	0.0185
218.0814032	3-Methylindolepyruvate	C12H11NO3	1.638958847	0.849090909	0.01759

Table 3. Continued...

M/Z	Metabolite	Molecular Formula	VIP	Fold change	P-value
191.1067823	{(1R,2R)-2-[(2Z)-5-Hydroxy-2-penten-1-yl]-3-oxocyclopentylacetic acid	C12H18O4	1.279848254	0.923436231	0.01046
195.1382105	Neocnidilide	C12H18O2	1.04252336	0.947191389	0.04929
279.1026762	Desmosflavone	C18H16O4	1.718820716	0.852964119	0.02664
206.1543316	Dihydrodioscorine	C13H21NO2	1.561470525	0.816051136	0.003173
222.112814	4-Hydroxy-2,6,6-trimethyl-3-oxo-1,4-cyclohexadiene-1-carboxaldehyde	C10H12O3	2.038780349	0.73527016	0.01187
135.0442898	R(-)-Mandelic acid	C8H8O3	1.923130401	0.731014493	0.018
339.0653247	5,7-Dihydroxy-4'-methoxy-8-methylflavanone	C17H16O5	1.581524301	0.811641067	0.04435
334.1556317	Phenylalanyl-Tryptophan	C20H21N3O3	2.763379521	0.655504395	0.004053
209.1540027	Riesling acetal	C13H22O3	1.220433156	0.875230486	0.04123
137.0963003	2-(1-Pentenyl)furan	C9H12O	1.204867687	0.89275766	0.02986
160.0250835	(E)-Raphanusanin	C6H9NS2	2.667747815	0.599025255	0.02478
132.0809797	3-Methylindole	C9H9N	1.990898857	0.817406909	0.002269
205.0974937	L-Tryptophan	C11H12N2O2	2.490258261	0.661298838	0.001893
215.081986	4-Quinolinecarboxylic acid	C10H7NO2	2.482321265	0.590557482	0.01978
249.0697831	Tyrosyl-Cysteine	C12H16N2O4S	1.522836797	0.881093936	0.004528
556.1069904	Dermatan	C18H31NO14S	1.336624727	0.904897379	0.03165
226.0866619	1-Hydroxy-10-methylacridone	C14H11NO2	1.247164667	0.892188336	0.03238
320.1170212	Epsilon-(gamma-Glutamyl)-lysine	C11H21N3O5	1.114360557	0.886113886	0.04471
104.0533741	2-Methylthiazolidine	C4H9NS	2.262661514	0.712335693	0.006966
207.0593045	Indolylmethylthiohydroximate	C10H10N2OS	1.547536814	0.853518574	0.001266
192.0659442	2-[[hydroxy(2-methoxyphenyl)methylidene]amino]acetic acid	C10H11NO4	1.464296912	0.859460809	0.02836
551.3675514	(3b,16a,21b,22a)-12-Oleanene-3,16,21,23,28-pentol-22-angeloyloxy-23-al	C35H54O7	2.273475675	0.701647875	0.03017
318.1567734	2-(3-Hydroxyphenyl)ethanol 1'-glucoside	C14H20O7	1.670972443	0.885862069	0.01386
146.0602316	Indole-3-carboxaldehyde	C9H7NO	1.44633741	0.886054096	0.04938
346.069578	Cochliophilin A	C16H10O5	2.007206951	0.758194655	0.01371
164.0469473	(S)-Isowillardine	C7H9N3O4	2.609031891	0.55723305	0.01135
170.060175	N-Cinnamoylglycine	C11H11NO3	2.461140769	0.43032049	0.03156
263.1395153	L-phenylalanyl-L-proline	C14H18N2O3	2.515837048	0.682747343	0.01846
136.0759556	2-Phenylacetamide	C8H9NO	1.212987977	0.915917991	0.0181
158.0813124	2-Hydroxyethylclavam	C7H11NO3	1.584284366	0.85918619	0.006938
215.1028348	N2-Succinyl-L-ornithine	C9H16N2O5	1.86050467	0.714202391	0.04268
292.0788124	Linamarin	C10H17NO6	2.047684699	0.765077821	0.01037
117.0736955	2-Methyl-1-methylthio-2-butene	C6H12S	2.157524575	0.759408912	0.00655
163.9837063	Raphanusamic acid	C4H5NO2S2	2.884861381	0.608154944	0.009327
520.2635865	11-Hydroxyprogesterone 11-glucuronide	C27H38O9	2.21978892	0.604443162	0.03875
117.0550738	Alpha-ketoisovaleric acid	C5H8O3	1.554512262	0.847489738	0.001027
263.0523276	Benzoyl glucuronide (Benzoic acid)	C13H14O8	2.15676507	0.712403323	0.004456
180.0657319	Adrenochrome	C9H9NO3	2.069599505	0.78973393	0.002385
139.0867471	2,4-Diaminoanisole	C7H10N2O	1.760802856	0.851221466	0.01367
297.0399462	3-(2,5-dihydroxy-4-methoxyphenyl)-6,7,8-trihydroxy-4H-chromen-4-one	C16H12O8	1.87701834	0.7583399	0.03086
89.0425623	Cis-2,3-Dimethylthiirane	C4H8S	1.89277777	0.808803913	0.01501
204.0506068	3-Fumarylpyruvate	C7H6O6	2.650917572	0.785977181	6.96E-05
117.0549626	2methyl-3-ketovaleric acid	C5H8O3	1.238531143	0.909182644	0.03561
203.1504826	ADMA	C8H18N4O2	1.272351038	0.409230769	0.03835
165.0758211	11-Hydroxyprogesterone 11-glucuronide	C6H14O6	2.029510943	0.655991165	0.04619
132.0656573	Trans-4-Hydroxy-L-proline	C5H9NO3	1.38260063	0.740924913	5.14E-05
689.2122744	Stachyose	C24H42O21	1.230116853	0.808257638	0.03233
665.5842526	9(R)-HODE cholesteryl ester	C45H76O3	2.604224801	0.395860495	0.001219
118.0863816	Betaine	C5H11NO2	1.161890218	0.90842442	0.00121

Table 3. Continued...

M/Z	Metabolite	Molecular Formula	VIP	Fold change	P-value
330.0935837	Amoxicillin	C16H19N3O5S	1.238582843	0.767499437	0.00447
144.0807517	2-Naphthylamine	C10H9N	1.677289415	0.735265431	0.009176
188.0709129	3-amino-2-naphthoic acid	C11H9NO2	1.495827215	0.743767573	0.008582
395.2911367	7,8-Dehydro-beta-microperoxanthin	C27H38O2	1.32836933	0.848883049	0.000134
487.3428115	Alisol C	C30H46O5	1.232155679	0.72293578	0.02673
407.219997	12alpha-Hydroxy-3-oxochola-1,4,6-trien-24-oic Acid	C24H32O4	1.848450052	0.686528497	6.46E-05
215.1394345	D-Dethiobiotin	C10H18N2O3	1.693484851	0.667572464	0.0001254
121.065066	P-Tolualdehyde	C8H8O	1.992266916	0.733934283	7.52E-05
162.076264	Pyroglutamic acid	C5H7NO3	1.875375131	0.528041181	0.001026
176.0918386	N-Carboxyethyl-gamma-aminobutyric acid	C7H13NO4	1.381360273	0.830349229	0.0009188
222.1338658	N-lactoyl-Valine	C8H15NO4	1.461137724	0.742272727	0.01644
118.086576	L-Valine	C5H11NO2	1.185474029	0.887396226	0.005473
220.118182	D-Pantothenic acid	C9H17NO5	1.333382266	0.81716663	3.42E-06
208.133505	(1xi,4xi,6xi)-Carvone oxide	C10H14O2	2.687298601	0.060368544	0.0007359
190.1075753	N-Methylcalystegine B2	C8H15NO4	1.029729661	0.780383186	0.009808
246.1703217	2-Methylbutyroylcarnitine	C12H23NO4	1.005160282	0.817808571	0.04581
232.154797	Isobutyryl carnitine	C11H21NO4	1.285682185	0.795703946	0.0007375
221.0808082	2,6-Dioxo-6-phenylhexanoate	C12H12O4	1.381787356	0.599235519	0.02316
393.2754149	3-Hydroxy-10'-apo-b,y-carotenal	C27H36O2	1.413759953	0.814967259	0.000119
426.1940115	Kanzonol N	C22H24O6	1.091618058	0.871246246	0.008901
377.2805347	10'-apo-beta-carotenal	C27H36O	1.316525996	0.850334305	0.0001803
399.3226623	1alpha,25-dihydroxy-9,11-didehydro-3-deoxyvitamin D3	C27H42O2	1.026653574	0.865434783	0.01144
375.2648843	Docosadienoate (22:2n6)	C22H40O2	1.626182871	0.69928667	0.0005543
415.3173691	Setariol	C27H42O3	1.41525567	0.782392386	0.0007405
174.0406958	4,5-Dihydro-5-methyl-3-thiophenethiol	C5H8S2	1.004974844	0.810749186	0.01591
226.1077942	L-4-Hydroxy-3-methoxy-a-methylphenylalanine	C11H15NO4	1.096440079	0.750062988	0.01625
208.0971492	Rhexifoline	C11H13NO3	1.150060485	0.710149166	0.01494
192.1021982	4-Ethoxybenzaldehyde	C9H10O2	1.391284686	0.718801539	0.001974
173.1075544	NORELEAGNINE	C11H12N2	1.524032747	0.672031212	0.003183
171.0918769	2-Amino-5-phenylpyridine	C11H10N2	1.387092349	0.779584277	7.47E-05
180.1021615	N-Acetyltyramine	C10H13NO2	1.191108945	0.819741481	0.003837
232.1548149	Butyryl-L-carnitine	C11H21NO4	1.715785742	0.47408051	0.02834
202.058461	Isocolumbin	C20H22O6	1.152282484	0.780523256	0.01246
251.1394956	Caffeoylputrescine	C13H18N2O3	1.409994252	0.680879121	0.02955
148.0758501	INDOLE-3-CARBINOL	C9H9NO	1.115604858	0.796278626	0.02604
220.1183375	Pantothenic Acid	C9H17NO5	1.642845241	0.725228272	0.004396
189.0073658	Alpha-Fluoro-beta-ureidopropionic acid	C4H7FN2O3	1.465630055	0.721371709	0.002484
77.06040416	1,3-Propanediol	C3H8O2	1.582739284	0.683584264	0.0001784
527.1594841	1-Kestose	C18H32O16	1.55501058	0.701506245	0.02238
188.1282478	3-hydroxyheptanoic acid	C7H14O3	1.671833191	0.698321649	0.0005467
220.1181122	2-O-Methyl-L-fucose	C7H14O5	1.749895787	0.702608696	2.44E-05
101.0600024	2-Ethylacrylic acid	C5H8O2	1.806588782	0.750922213	1.04E-06
188.1758827	N1-Acetylspermidine	C9H21N3O	1.617652422	0.774978578	0.002106
160.1333685	METHACHOLINE	C8H17NO2	1.691271785	0.677038109	4.20E-06
263.2372177	Linoleic acid	C18H32O2	1.352169651	0.851196329	0.03838
132.0656573	Trans-4-Hydroxy-L-proline	C5H9NO3	1.867926684	0.692057594	0.001464
565.1564416	Apiin	C26H28O14	2.097536921	0.660624551	0.01451
369.1727038	Quercetol C	C22H24O5	1.016354691	0.907909605	0.01797
671.3780223	Phytolaccoside A	C36H56O10	1.395740126	0.833242009	0.03876
437.1272588	Para-hydroxyrosiglitazone	C18H19N3O4S	1.690840196	0.729662361	0.006775
446.1641876	Neolinustatin	C17H29NO11	1.625058167	0.654886803	0.02886
487.3428115	Alisol C	C30H46O5	1.618683744	0.741220736	0.01598

Table 3. Continued...

M/Z	Metabolite	Molecular Formula	VIP	Fold change	P-value
137.0598335	Benzylformate	C8H8O2	1.549603749	0.803608737	0.01765
407.219997	12alpha-Hydroxy-3-oxochola-1,4,6-trien-24-oic Acid	C24H32O4	2.288070638	0.695652174	0.0004605
125.0599248	P-Methylolphenol	C7H8O2	1.14943481	0.86203847	0.03443
191.1069182	(R)-3,7-Dimethyl-5-indanecarboxylic acid	C12H14O2	1.195838985	0.825458052	0.03843
75.05597202	N-Methylurea	C2H6N2O	1.254449906	0.827370304	0.01523
281.08029	Gamma-Glutamyl-S-methylcysteine sulfoxide	C9H16N2O6S	1.647287262	0.754361339	0.01396
222.1338658	N-lactoyl-Valine	C8H15NO4	1.46940662	0.762727697	0.02713
220.118182	D-Pantothenic acid	C9H17NO5	1.271832463	0.867503487	0.0001832
291.0392706	N5-(3,4-Dioxo-1,5-cyclohexadien-1-yl)-L-glutamine	C11H12N2O5	1.850808617	0.588296761	0.01681
209.1175726	(4-tert-Butyl-phenoxy)-acetic acid	C12H16O3	1.603682692	0.779003088	0.003158
217.0498642	ISOBERGAPTENE	C12H8O4	2.684947086	0.422617843	0.004417
218.0814032	3-Methylindolepyruvate	C12H11NO3	1.252178535	0.852727273	0.007778
432.2727937	Gamma-Eudesmol rhamnoside	C21H36O5	1.135897354	0.857176526	0.03914
418.2568652	4-Hydroxy-D4-neuroprostane	C22H32O5	1.2734631	0.842557652	0.0412
376.2463469	(ent-16betaOH)-16,17-Dihydroxy-9(11)-kauren-19-oic acid	C20H30O4	1.439696342	0.774355301	0.03734
195.1382105	Neocnidilide	C12H18O2	1.012091056	0.925832492	0.01384
445.3139915	Hydroxysintaxanthin 5,6-epoxide	C31H42O3	1.184997554	0.84363879	0.02443
396.2727566	PGE2-EA	C22H37NO5	1.246325863	0.852593842	0.03367
C18H16O4	Desmosflavone	C18H16O4	1.373241238	0.840678627	0.01035
255.0655761	Daidzein	C15H10O4	1.204115555	0.892776755	0.0006034
223.0602868	3-(6-hydroxy-7-methoxy-2H-1,3-benzodioxol-5-yl)prop-2-enal	C11H10O5	1.108424445	0.877188594	0.003953
104.0533741	2-Methylthiazolidine	C4H9NS	1.446386897	0.760616785	0.02352
214.0633156	Cysteiny-Glutamine	C8H15N3O4S	1.312993536	0.873096447	0.001057
195.0879892	Caffeine	C8H10N4O2	1.576412125	0.570684039	0.03142
195.1130541	4,5-Dimethylxazole	C5H7NO	1.047886711	0.89216152	0.003581
202.058461	Isocolumbin	C20H22O6	1.101813786	0.855776892	0.02373
227.1030103	Porphobilinogen	C10H14N2O4	1.071135255	0.891752577	0.03301
160.1333685	METHACHOLINE	C8H17NO2	1.511227935	0.737907466	0.02717
77.06040416	1,3-Propanediol	C3H8O2	1.515237808	0.728142783	0.0174
117.0549626	2methyl-3-ketovaleric acid	C5H8O3	1.407986888	0.845786963	0.0009994
527.1594841	1-Kestose	C18H32O16	1.358205407	0.790846966	0.04887
141.018272	Cis-4-Carboxymethylenebut-2-en-4-olide	C6H4O4	1.175827063	0.885444744	0.002208
101.0600024	2-Ethylacrylic acid	C5H8O2	1.685010841	0.799658412	0.002509

Shaanxi. It has high nutritional value and the fermentation process is simple. The functional compound probiotics in the serous water can directly or indirectly affect the healthy microorganisms in the intestinal system, showing its important role in the human gastrointestinal system (Kavas et al., 2022). However, the microbial community structure of Jiangshui is complex (Chen et al., 2016b). In this study, high-throughput sequencing technology was used to analyze the microbial community structure and diversity in Jiangshui from different regions, explore the differences between the dominant bacterial groups in different regions, and determine the flavor differences in Jiangshui from different regions based on metabolomics technology, to further improve flavor quality and nutritional value.

The abundance and diversity of microbial communities in Jiangshui samples changed in different regional environments (Fang et al., 2015). The sequencing depth index was > 99% and

the species accumulation curve showed the Jiangshui sequencing samples and sequencing data were reasonable. *Lactobacillus* was the dominant strain of the bacterial community at the species level (Liu et al., 2018), and the fungal community structure showed diversity. The laboratory strains isolated by Li et al. (2021) from the Jiangshui samples from Tianshui, Gansu, Northwest China, mainly belonged to *Lactobacillus* and *Bifidobacterium* (Li et al., 2021). Liang et al. (2018) also found that lactic acid bacteria were dominant in Northeast sauerkraut and Sichuan pickle. Li et al. (2022) analyzed the microbial structure of Jiangshui from different regions using MIDI MIS and found the proportion of bacteria and fungi was as high as 93.05% and both play an important role in the formation of Jiangshui flavor and improvement of nutritional effect. Tadjine et al. (2021) found that cheese made from raw milk contains a variety of lactic acid bacteria, such as *Lactococcus*, *Lactobacillus*, *Leuconostoc* and *Enterococcus* (Tadjine et al., 2021). Therefore, although

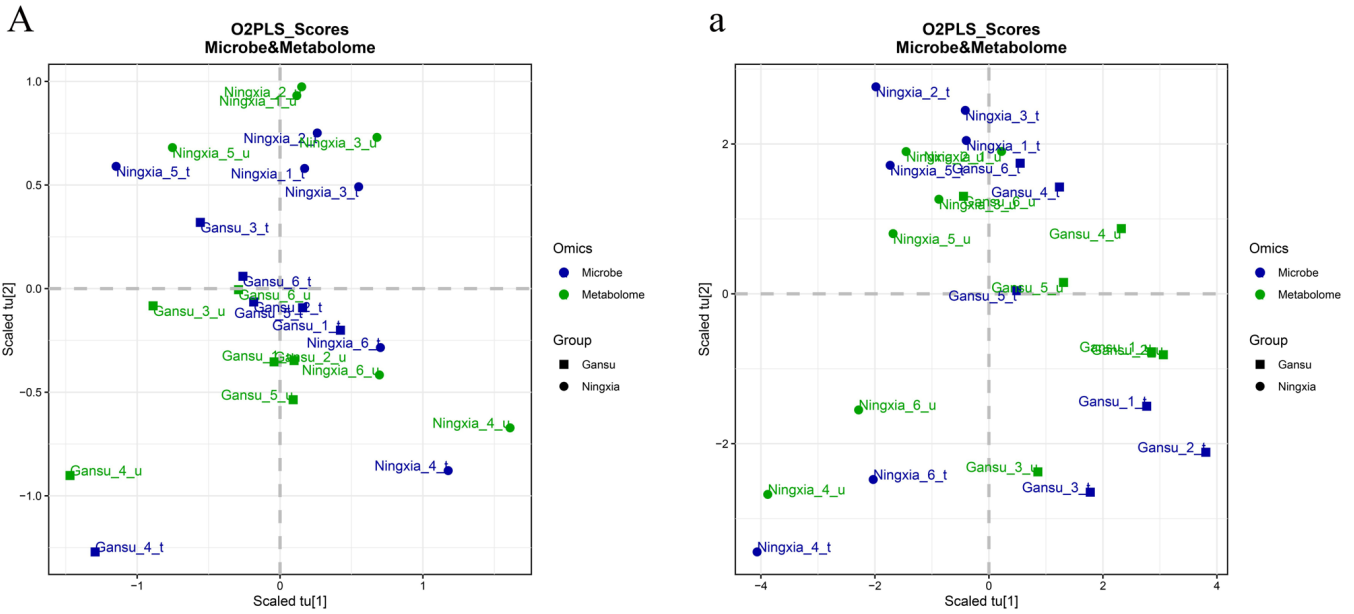


Figure 11. O2PLS (the two-way orthogonal partial least squares to assess the intrinsic correlation between the two data sets) score map of the different regions. The shapes (squares and dots) represent different groups of samples, the colors represent different omics samples, where blue represents microbiome samples. The sample name is followed by t. The green represents metabolome samples, and the sample name is followed by u. The abscissa and ordinate represent the scores of the combination of metabolome and microbiome, u represents the scores of microbial samples, and u represents the scores of metabolome samples. A: GS and NX bacterial community; a: GS and NX fungal communities, and microbial communities.

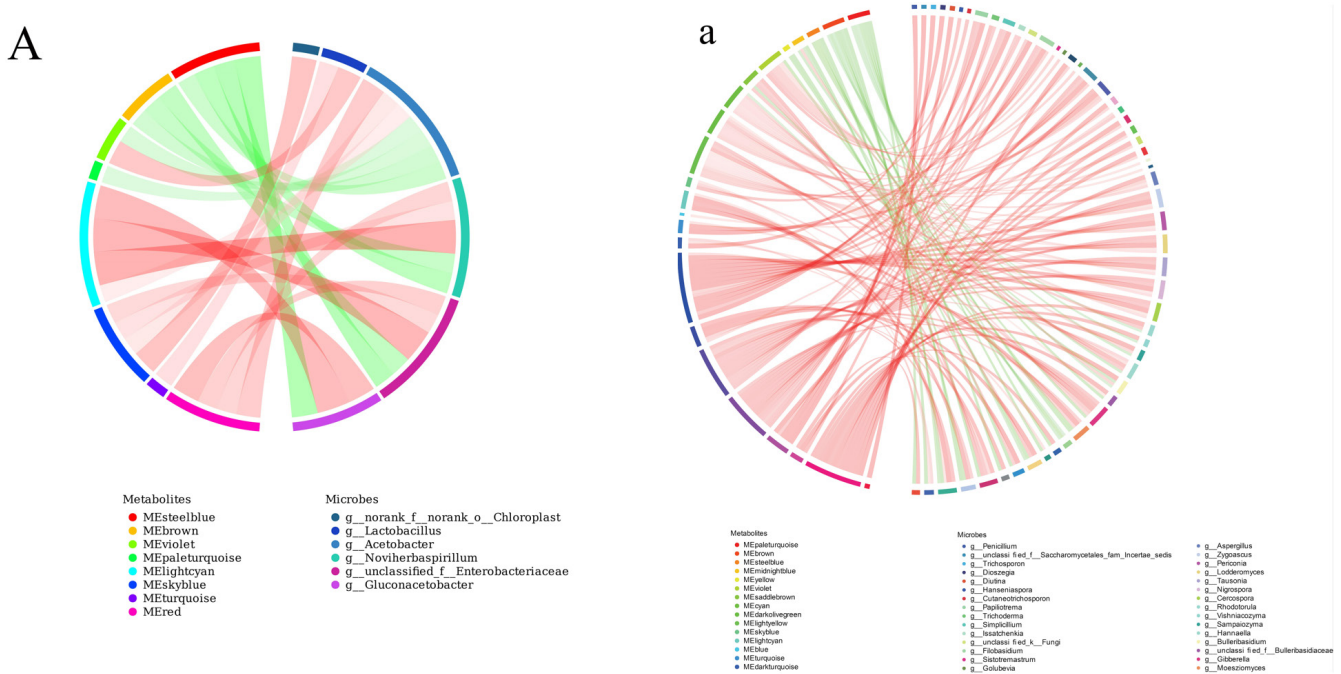


Figure 12. WGCNA (weighted gene co-expression network analysis) correlation chord diagram of Jiangshui from different regions was used to decompose the metabolites into different metabolite modules. The metabolites in the modules were highly correlated, and the correlation analysis with the flora was performed. The left half circle of the chord diagram is the metabolite and the right half circle is the microorganism. Each string indicates the metabolite has a significantly high correlation with the microorganism. The red string represents a positive correlation and the green string represents a negative correlation. The wider the width of the string, the more counts associated with this metabolite or microorganism. A: Correlation between GS and NX bacterial flora and metabolites; a: correlation between GS and NX fungal flora and metabolites, and the correlation between microbial communities and metabolites in other regions are shown.

differences in fermentation processes and regions may affect microbial community diversity in the Jiangshui, *Lactobacillus* is the main bacterial community that determines the changes in the bacterial community in Jiangshui.

In the present study, LC-MS untargeted metabolomics analysis was used to study the different metabolites in Jiangshui from three regions (Liu et al., 2021). Santos et al. (2021) also used LC-MS technology to analyze the mature fruits of EECC (Ethanol extract from *Capsicum chinense*) and identified 10 phenolic compounds (Santos et al., 2021). PCA showed that the distance between regions directly affected the aggregation degree of Jiangshui samples. The discriminant and permutation tests of OPLS-DA showed the model was highly reliable, improved the effectiveness and analytical ability of the model, could better distinguish the differences between groups, and could be used for subsequent differential metabolite analysis (Rao et al., 2016). Based on $VIP > 1$, $P\text{-value} < 0.05$, and $FC < 1$, the identification of differential metabolites and pathway enrichment analysis were performed. For example, compared to Jiangshui from Guyuan, Ningxia, Jiangshui from Tianshui, Gansu, had more metabolic pathways for protein digestion and absorption, aminoacyl-tRNA biosynthesis, central carbon metabolism in cancer, cyanoamino acid metabolism, biosynthesis of plant secondary metabolites, and cocaine addiction. There were seven differential metabolites enriched in the metabolic pathways of protein digestion and absorption and aminoacyl-tRNA biosynthesis, including L-tyrosine, L-threonine, L-valine, histidine, phenylalanine, L-tryptophan, and L-glutamate. These amino acids contribute to the main flavor of Jiangshui caused by metabolic utilization of Jiangshui in different regions. The same metabolites simultaneously participate in multiple metabolic pathways, indicating the different metabolites have a significant effect on the pathway. Various effects on the above metabolic pathways can change the content of metabolic substances in the Jiangshui from various regions as well as improve the taste quality, nutritional value, and efficacy of the Jiangshui. Finally, based on joint analysis of microbial diversity and metabolomics, the Jiangshui samples from the different regions were highly correlated (McHardy et al., 2013).

5 Conclusion

In the present study, 16S rRNA high-throughput sequencing and untargeted metabolomics were used to analyze the microbial community diversity and differential metabolites in 18 Jiangshui samples from Tianshui, Gansu, Guyuan, Ningxia, and Ankang, Shaanxi. The results showed several regional differences in microorganisms in Jiangshui from the three different areas. The bacterial community structure was relatively stable and diversity was low. *Lactobacillus* was the dominant genus in Jiangshui from the three regions; however, the fungal community structure was complex and changeable. Furthermore, 100 genera were detected in the Jiangshui samples from the three regions at the genus classification level. The dominant genera in the different regions had significant differences in species and abundance. Based on $VIP > 1$, $P < 0.05$, and $FC < 1$, 78 different metabolites were screened out in the Jiangshui from the three regions using ESI^+ . Among them, the key metabolites L-tyrosine, L-threonine,

L-valine, histidine, phenylalanine, L-tryptophan, and L-glutamate were enriched in protein digestion and absorption and biosynthesis of aminoacyl-tRNA. These differential metabolites formed a highly sensory pleasant flavor of the Jiangshui (Jun et al., 2018; Moon et al., 2018). Based on joint analysis, the microorganisms and metabolites in the Jiangshui from the three different regions correlated, and the metabolites/microorganisms were significantly associated with various microorganisms/metabolites. Therefore, our results clarify the differences between Jiangshui from different regions and provide a theoretical basis for future studies on the composition and content of metabolites in Jiangshui from other regions.

References

- Cao, J., Yang, J., Hou, Q., Xu, H., Zheng, Y., Zhang, H., & Zhang, L. (2017). Assessment of bacterial profiles in aged, home-made Sichuan paocai brine with varying titratable acidity by PacBio SMRT sequencing technology. *Food Control*, 78, 14-23. <http://dx.doi.org/10.1016/j.foodcont.2017.02.006>.
- Chen, P., Wu, Z., Zhao, Y., Wei, Y., Xu, R., Yan, L., & Li, H. (2016a). Cultivation-independent comprehensive investigations on bacterial communities in serofluid dish, a traditional Chinese fermented food. *Genomics Data*, 7, 127-128. <http://dx.doi.org/10.1016/j.gdata.2015.12.019>. PMID:26981386.
- Chen, P., Zhao, Y., Wu, Z., Liu, R., Xu, R., Yan, L., & Li, H. (2016b). Metagenomic data of fungal internal transcribed spacer from serofluid dish, a traditional Chinese fermented food. *Genomics Data*, 7, 134-136. <http://dx.doi.org/10.1016/j.gdata.2015.12.028>. PMID:26981389.
- Fang, R. S., Dong, Y. C., Chen, F., & Chen, Q. H. (2015). Bacterial diversity analysis during the fermentation processing of traditional Chinese yellow rice wine revealed by 16S rDNA 454 pyrosequencing. *Journal of Food Science*, 80(10), M2265-M2271. <http://dx.doi.org/10.1111/1750-3841.13018>. PMID:26409170.
- Hou, J. C., Jiang, C., & Long, Z. (2013). Nitrite level of pickled vegetables in Northeast China. *Food Control*, 29(1), 7-10. <http://dx.doi.org/10.1016/j.foodcont.2012.05.067>.
- Jun, Z., Shuaishuai, W., Lihua, Z., Qilong, M., Xi, L., Mengyang, N., Tong, Z., & Hongli, Z. (2018). Culture-dependent and -independent analysis of bacterial community structure in Jiangshui, a traditional Chinese fermented vegetable food. *LWT*, 96, 244-250. <http://dx.doi.org/10.1016/j.lwt.2018.05.038>.
- Kavas, N., Kavas, G., Kinik, Ö., Ateş, M., Kaplan, M., & Şatir, G. (2022). Symbiotic microencapsulation to enhance *Bifidobacterium longum* and *Lactobacillus paracasei* survival in goat cheese. *Food Science and Technology*, 42, e55620. <http://dx.doi.org/10.1590/fst.55620>.
- Li, P., Ju, N., Zhang, S., Wang, Y., & Luo, Y. (2022). Evaluation of microbial diversity of Jiangshui from the Ningxia Hui autonomous region in China. *Food Biotechnology*, 36(2), 173-190. <http://dx.doi.org/10.1080/08905436.2022.2054818>.
- Li, S., Deng, B., Tian, S., Guo, M., Liu, H., & Zhao, X. (2021). Metabolic and transcriptomic analyses reveal different metabolite biosynthesis profiles between leaf buds and mature leaves in *Ziziphus jujuba* mill. *Food Chemistry*, 347, 129005. <http://dx.doi.org/10.1016/j.foodchem.2021.129005>. PMID:33482487.
- Liang, H. P., Yin, L. G., Zhang, Y. H., Chang, C., & Zhang, W. X. (2018). Dynamics and diversity of a microbial community during the fermentation of industrialized Qingcai paocai, a traditional Chinese fermented vegetable food, as assessed by Illumina MiSeq sequencing, Dgge and qPCR assay. *Annals of Microbiology*, 68(2), 111-122. <http://dx.doi.org/10.1007/s13213-017-1321-z>.

- Liu, X. J., Zhou, M., Jiabin, C. X., Luo, Y., Ye, F. Z., Jiao, S., Hu, X. Z., Zhang, J. C., & Lu, X. (2018). Bacterial diversity in traditional sourdough from different regions in China. *Lebensmittel-Wissenschaft + Technologie*, 96, 251-259. <http://dx.doi.org/10.1016/j.lwt.2018.05.023>.
- Liu, X., Fan, H. M., Liu, D. H., Liu, J., Shen, Y., Zhang, J., Wei, J., & Wang, C. L. (2021). Transcriptome and metabolome analyses provide insights into the watercore disorder on "Akiba" pear fruit. *International Journal of Molecular Sciences*, 22(9), 4911. <http://dx.doi.org/10.3390/ijms22094911>. PMID:34066340.
- Lü, X., Yi, L. H., Dang, J., Dang, Y., & Liu, B. F. (2014). Purification of novel bacteriocin produced by *Lactobacillus coryniformis* MXJ 32 for inhibiting bacterial foodborne pathogens including antibiotic-resistant microorganisms. *Food Control*, 46, 264-271. <http://dx.doi.org/10.1016/j.foodcont.2014.05.028>.
- McHardy, I. H., Goudarzi, M., Tong, M., Ruegger, P. M., Schwager, E., Weger, J. R., Graeber, T. G., Sonnenburg, J. L., Horvath, S., Huttenhower, C., McGovern, D. P. B., Fornace, A. J. Jr., Borneman, J., & Braun, J. (2013). Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome*, 1(1), 17. <http://dx.doi.org/10.1186/2049-2618-1-17>. PMID:24450808.
- Moon, S. H., Kim, C. R., & Chang, H. C. (2018). Heterofermentative lactic acid bacteria as a starter culture to control kimchi fermentation. *Lebensmittel-Wissenschaft + Technologie*, 88, 181-188. <http://dx.doi.org/10.1016/j.lwt.2017.10.009>.
- Nguyen, D. T. L., Van Hoorde, K., Cnockaert, M., Brandt, E., Aerts, M., Thanh, L. B., & Vandamme, P. (2013). A description of the lactic acid bacteria microbiota associated with the production of traditional fermented vegetables in Vietnam. *International Journal of Food Microbiology*, 163(1), 19-27. <http://dx.doi.org/10.1016/j.ijfoodmicro.2013.01.024>. PMID:23500611.
- Rao, G., Sui, J., & Zhang, J. (2016). Metabolomics reveals significant variations in metabolites and correlations regarding the maturation of walnuts (*Juglans regia* L.). *Biology Open*, 5(6), 829-836. <http://dx.doi.org/10.1242/bio.017863>. PMID:27215321.
- Santos, L. S., Fernandes, C. C., Santos, L. S., Deus, I. P. B., Sousa, T. L., & Miranda, M. L. D. (2021). Ethanolic extract from *Capsicum chinense* Jacq. ripe fruits: phenolic compounds, antioxidant activity and development of biodegradable films. *Food Science and Technology*, 41(2), 497-504. <http://dx.doi.org/10.1590/fst.08220>.
- Tadjine, D., Boudalia, S., Bousbia, A., Gueroui, Y., Symeon, G., Boudechiche, L. M., Tadjine, A., & Chemmam, M. (2021). Milk heat treatment affects microbial characteristics of cows' and goats' "Jben" traditional fresh cheeses. *Food Science and Technology*, 41(1), 136-143. <http://dx.doi.org/10.1590/fst.00620>.
- Wang, H., Wu, Y., Xiang, H., Sun-Waterhouse, D., Zhao, Y., Chen, S., Li, L., & Wang, Y. (2022). UHPLC-Q-Exactive Orbitrap MS/MS-based untargeted lipidomics reveals molecular mechanisms and metabolic pathways of lipid changes during golden pomfret (*Trachinotus ovatus*) fermentation. *Food Chemistry*, 396, 133676. <http://dx.doi.org/10.1016/j.foodchem.2022.133676>. PMID:35868287.
- Wang, Y., Li, C., Li, L., Yang, X., Chen, S., Wu, Y., Zhao, Y., Wang, J., Wei, Y., & Yang, D. (2019). Application of UHPLC-Q/TOF-MS-based metabolomics in the evaluation of metabolites and taste quality of Chinese fish sauce (Yu-lu) during fermentation. *Food Chemistry*, 296, 132-141. <http://dx.doi.org/10.1016/j.foodchem.2019.05.043>. PMID:31202297.
- Zhang, Q., Wang, C., Li, B., Li, L., Lin, D., Chen, H., Liu, Y., Li, S., Qin, W., Liu, J., Liu, W., & Yang, W. (2018). Research progress in tofu processing: from raw materials to processing conditions. *Critical Reviews in Food Science and Nutrition*, 58(9), 1448-1467. <http://dx.doi.org/10.1080/10408398.2016.1263823>. PMID:27977295.

Supplementary Material

Supplementary material accompanies this paper.

Figure S1. PCA scores of Jiangshui samples from different regions using ESI⁻.

Figure S2. (A–C) are the OPLS-DA scores using ESI⁻, the OPLS-DA score map filters out the information irrelevant to the group through orthogonal rotation to better distinguish the differences between groups and improve the efficiency of the model. A and a: GS vs. NX; B and b: GS vs. SX; C and c: NX vs. SX.

Figure S3. ESI⁻ volcano map of the difference between the Jiangshui groups from different regions.

Figure S4. O2PLS score map of the different regions. B: SX and GS bacterial communities; b: SX and GS fungal community, C: SX and GS bacterial community; c: SX and GS fungal communities.

Figure S5. The WGCNA correlation chord diagram of Jiangshui from different regions used to decompose the metabolites into different metabolite modules. B: Correlation between SX and GS bacterial flora and metabolites; b: Correlation between SX and GS fungal flora and metabolites, C: Correlation between SX and NX bacterial flora and metabolites; c: Correlation between SX and NX fungal flora and metabolites.

Table S1. Differences in metabolites in Jiangshui samples from different regions based on ESI⁻.

This material is available as part of the online article from <https://doi.org/10.1590/fst.107222>