Current Research in Food Science 11 (2025) 101189

Contents lists available at ScienceDirect

Current Research in Food Science

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

Metagenomic insights into viral dynamics and funcation in Baijiu

Huadong Zhang [a], Hongxia Zhang [b], Hai Du [a], Yan Zhang [c], Menghui Zhang [d], Xiaowei Yu [a,*], Yan Xu [a,**]

[a] Laboratory of Brewing Microbiology and Applied Enzymology, The Key Laboratory of Industrial Biotechnology, Ministry of Education, State Key Laboratory of Food Science and Technology, School of Biotechnology, Jiangnan University, 1800 Lihu Ave, Wuxi, 214122, Jiangsu, China
[b] College of Life Sciences, Shanxi Normal University, Taiyuan, 030000, Shanxi, China
[c] Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China
[d] State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Keywords:
Baijiu fermentation
virus
phage
Metagenomic analysis
Microbial community

ABSTRACT

Baijiu fermentation represents a sophisticated microbial-driven biochemical process mediated by complex microbial consortium. Despite extensive characterization of bacterial and fungal roles in fermentation systems, the virus remains a critical knowledge gap. In this study,we employed metagenomics to profile the dynamics of viral community in fermented grains across five stages (day 0, 5, 10, 20 and 30) of a Baijiu fermentation. The metagenomics revealed 101 viral families, dominated by Metaviridae, Parvoviridae, Aliceevansviridae, Herelleviridae, Geminiviridae, Iridoviridae, and Genomoviridae, and lactic acid bacteria was identified as primary phage hosts. The results revealed that ssRNA viruses and ssDNA viruses were more abundant during 0–5 days, dsDNA viruses became dominant during 10–30 days. Multivariate analysis indicated that the viral community dynamics during the fermentation were primarily governed by microbes in succession, environmental factors (temperature,pH, moisture and glucose) and metabolites (lactic acid, acetate and ethanol) in the biosystem. Notably, predicting phages exhibited strong positive correlations with their respective hosts ($P < 0.01$, $r > 0.6$). We have identified viral auxiliary metabolic genes (AMGs) related to amino acid metabolism and vitamin biosynthesis. At 30 days of fermentation, the number and abundance of AMGs significantly increased. Our findings provide novel insights into the viral ecology in complex Baijiu fermentation ecosystem, shedding light on the intricate interactions within fermentation microbial communities.

1. Introduction

Baijiu is produced through a traditional spontaneous fermentation process that relies on the metabolism and succession of multiple microorganisms (Huang et al., 2025; Wang et al., 2020). During the pre-fermentation phase, aerobic microorganisms, such as molds and *Bacillus* spp. dominate, producing glucoamylase and proteases that break down macromolecules in the grain (Wu et al., 2021). This process converts these macromolecules into carbon and nitrogen sources, which can be utilized by other microorganisms during subsequent stages of fermentation (Wei et al., 2023). As the oxygen concentration in the cell decreases, yeasts begin to dominate, producing ethanol and flavor compounds (Luo et al., 2023; You et al., 2021). Subsequently, as ethanol concentration increases, the numbers of various alcohol-intolerant microorganisms gradually decline (Cai et al., 2022). In the later stages of fermentation, lactic acid bacteria become the dominant group, continuing to produce lactic acid and flavor compounds (Hu et al., 2021). The correct succession of microorganisms and their interactions are essential to ensure the production of high-quality Baijiu (Tan et al., 2019). However, in contrast to the well-documented and rich succession patterns of bacteria and fungi, our understanding of virus within the Baijiu fermentation process remains limited and largely uncharted.

In past research on fermented food systems, viruses —particularly bacterial viruses (phages)—have received the most attention in the field of virology. Phages are directly associated with host physiology and mortality by virulent and temperate lifestyles, and further influence microbial community dynamics in fermented food systems (Zhang et al., 2024). For example, the phage community exhibits a significant negative correlation with the bacterial community in terms of alpha diversity during the fermentation of traditional rice vinegars. Additionally,

* Corresponding author.
** Corresponding author.
E-mail addresses: yuxw@jiangnan.edu.cn (X. Yu), yxu@jiangnan.edu.cn (Y. Xu).

https://doi.org/10.1016/j.crfs.2025.101189
Received 2 June 2025; Received in revised form 4 August 2025; Accepted 29 August 2025
Available online 7 September 2025

phages play a unique role in rice vinegar fermentation by expressing numerous carbohydrate auxiliary metabolic genes (AMGs) (Ma et al., 2024). Several viral AMGs associated with acetate synthesis have been identified in grain vinegar fermentation (Yu et al., 2022). Moreover, there are significant differences in the viral community in Daqu at different temperatures, and correlations have been observed between the viral and bacterial communities (Huang et al., 2025; Kang et al., 2022). Additionally, during the fermentation of sauce-flavored Baijiu, phage induction liquids have a significant inhibitory effect on *Bacillus* spp. (Du et al., 2022). In the fermentation of strong-flavor Baijiu, Spearman's rank correlation analysis revealed a negative correlation between Peduoviridae and *Gluconobacter*. Genomoviridae also shows negative correlations with *Issatchenkia*, *Penicillium*, and *Monascus* (Zhang et al., 2025).

With most recent high-throughput sequencing technology, metagenomics and viral metagenomics have become powerful tools for exploring viral diversity and function (Paez-Espino et al., 2016; Sommers et al., 2021), each has its own merits and disadvantages. Metagenomics uses unbiased sequencing to simultaneously detect bacterial, fungal, viral communities and even prophages integrated within bacterial genomes in a biosystem (Liu et al., 2022). This method might ask for higher sequencing depth in observation of virus since it sequences all microbial cellular DNA in a biosystem and viral sequences typically constitute a relatively small fraction (Trubl et al., 2020). The viral metagenomics involves a viral particle enrichment process that specifically targets and sequences viral DNA, significantly increasing the proportion of viral sequences (Wang et al., 2023). However, this enrichment process inadvertently leads loss of some viruses and introduces bias in the amplification of certain viral DNA (Callanan et al., 2021). In a previous study, the metagenomics approach detected a greater number of viral sequences and AMGs in wastewater, whereas the viromic approach yielded higher-quality viral sequences and identified more eukaryotic viruses (Zhang et al., 2024). In soil virus research, researchers have found that compared to metagenomics, the viromic approach can detect a broader and more comprehensive range of viral communities, facilitating the discovery of new viruses (Santos-Medellin et al., 2021). Viromic methods have been primarily used in previous studies on Baijiu viruses (Du et al., 2022; Huang et al., 2025; Kang et al., 2022; Zhang et al., 2025). However, the advantages and limitations of metagenomics in detecting viruses during Baijiu fermentation remain unclear.

In this study, we employed metagenomics approaches to characterize the viral community dynamics in fermented grains during the process of a strong-flavor Baijiu fermentation. We examined the distribution of host-associated genes and AMGs among the identified viruses. Our results showed the succession characteristics of the viral community and its potential auxiliary metabolic gene function.

2. Materials and methods

2.1. Sample collection

To compare the viral communities in fermented grains at different fermentation stages, we collected samples in fermented grains on day 0, 5, 10, 20, and 30 from a well-known strong-flavor Baijiu producing pipeline (Fig. 1A) in Sichuan Province, China, in January 2023, based on previous research (Du et al., 2022; Yuan et al., 2023). These five time points correspond to the three stages outlined in the introduction: 0–10 days (pre-fermentation), 10–20 days (mid-fermentation), and 20–30 days (post-fermentation). Fermentation tracking was conducted in three selected fermentation mud pits under normal fermentation conditions. Each fermented grain sample was collected from three different depths (0.5, 1, and 1.5 m) (as shown in Fig. 1B and C). Approximately 100 g of material was collected per sample, yielding a total of 15 samples. All samples were rapidly frozen in liquid nitrogen and stored in the laboratory at −80 °C until analysis.

2.2. Total microbial cell acquisition

For the traditional metagenomics sequencing (Song et al., 2019), approximately 5 g of each sample was added to 30 mL of PBS buffer, followed by vortexing and shaking for 5 min. Microbial cells were then collected through gradient centrifugation. Initially, the mixture was centrifuged at 30×*g* for 10 min, after which the supernatant was subjected to a second centrifugation at 2000×*g* for 10 min to isolate the total microbial cell fraction.

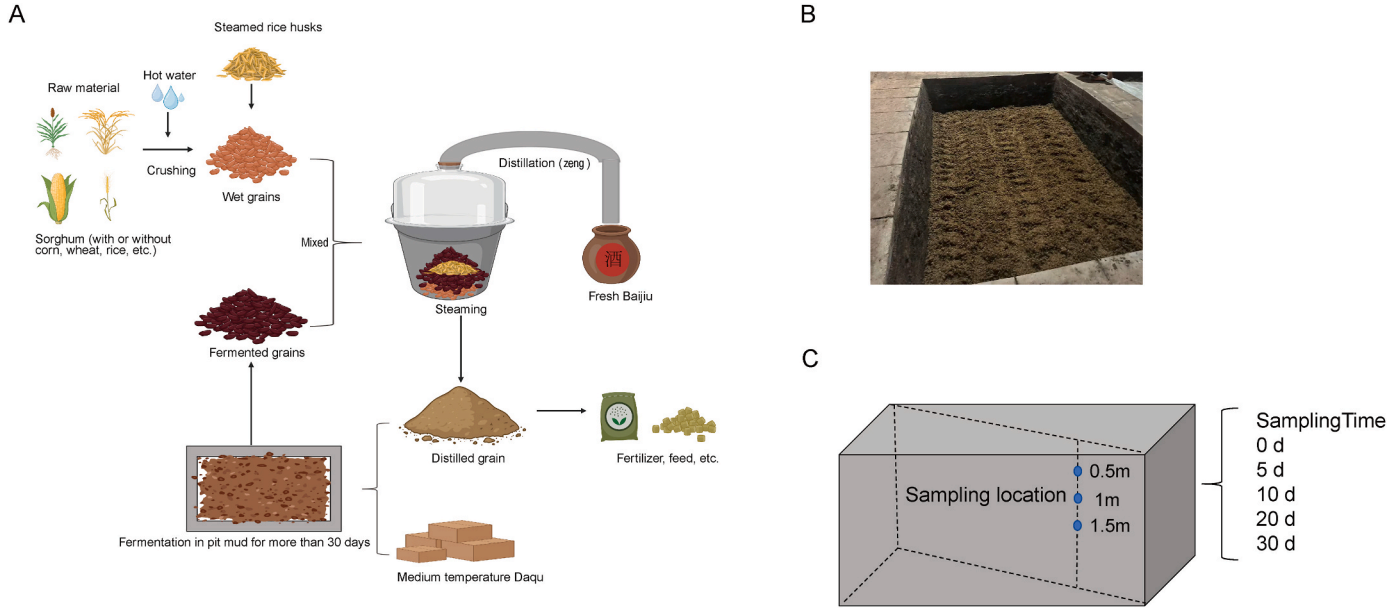

Fig. 1. Schema of traditional repeated-batch process of strong-flavor Baijiu and sampling details. (A) Strong-flavor Baijiu production process. (B) Actual view of the sampling plant. (C) Fermentation tracking was conducted in selected fermentation mud pits under normal fermentation conditions. Fermented grain samples were collected from three different depths (0.5, 1, and 1.5 m).

2.3. Nucleic acid extraction and sequencing

The total microbial DNA from the sample was extracted using the FastDNA™ Spin Kit for Soil (MP Biomedicals). Metagenomics sequencing libraries were constructed using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs). Bipartite 150 bp sequencing reads were generated on the Illumina NovaSeq 6000 platform (performed by Megger Biotechnology Ltd.). The total microbial DNA were sequenced, generating approximately 10 GB of raw data.

The raw data were processed using Trimmomatic (v0.39) with default parameters to obtain clean data for subsequent analysis (Bolger et al., 2014). Afterwards, the contigs from the bacterial, fungal, and viral fractions were assembled separately using Megahit (v1.1.2) with the following parameters: k-min 35, k-max 95, and k-step 20 for the bacterial and fungal fractions, and –presets meta-large, –min-contig-len 300 for the viral fraction (Li et al., 2015).

2.4. Bacterial and fungal taxonomic assignment and functional annotation

Open reading frames (ORFs) were predicted using Prodigal (V2.6.3) (Hyatt et al., 2010) with the parameters -f gff, -p meta, and -p 11. Redundancy was then eliminated using MMseqs2 (Steinegger and Söding, 2017) based on 95 % sequence similarity and 90 % coverage to obtain the unigenes catalog. Then, the gene catalogs were mapped to the clean data using BBMap (v38.90, https://github.com/BioInfoTools/BBMap) with default parameters to determine the abundance of genes in each sample. Bacterial and fungal taxonomy was further annotated using the lowest common ancestor (LCA) algorithm in MEGAN 5 (Huson et al., 2011), with the NCBI-NT reference database accessed through BLASTn. Functional gene annotation was performed by aligning sequences to the KEGG and eggNOG (v4.5.1) databases using DIAMOND (v0.9.32.133) (Buchfink et al., 2015) with an e-value threshold of $\leq$0.001.

2.5. Viral contig identification and quantification

Since viruses lack marker genes like those found in bacteria and fungi, we used four tools to improve the accuracy of virus recognition and identify candidate viral contigs: geNomad (v1.8.0) (Camargo et al., 2023), Virsorter2 (v2.2.4) (Guo et al., 2021), VIBRANT (v1.2.1) (Kieft et al., 2020), and viralVerify (v1.1) (Antipov et al., 2020). Each tool was employed with its default settings to screen for viruses. The viruses identified by these tools were then merged to create a set of non-redundant, unique candidate viral contigs. This ensemble of candidate contigs was subsequently analyzed. The quality and completeness of candidate viral contigs were assessed using CheckV (v1.0.3) (Nayfach et al., 2020). This tool improves viral sequence identification accuracy by screening contigs for viral proteins and retaining only those $\geq$2 kb in length that contain at least one viral protein hit for downstream analysis (Cantuti Gendre et al., 2025; Zhang et al., 2024). Filtered viral contigs were dereplicated and clustered into viral operational taxonomic units (vOTUs) using CD-HIT (v4.8.1) with the parameters: "-c 0.95 -n 5" (Fu et al., 2012). The relative abundance of vOTUs in each sample was quantified as RPKM values. These values were calculated using CoverM (v0.6.1) (https://github.com/wwood/CoverM) with the parameters: "–contig –min-read-aligned-percent 75 –min-read-percent-identity 95 –min-covered-fraction 75 –contig-end-exclusion 0 -m rpkm." The alpha diversity indices of the viral community were calculated using R (v4.3.0). Non-metric multidimensional scaling analysis (NMDS) based on Bray–Curtis distances was performed to assess community dissimilarities.

2.6. Viral contig taxonomic assignment, host analysis and functional annotation

The taxonomic classification of viral contigs was conducted using two methods, following the latest classification rules of the International Committee on Taxonomy of Viruses (ICTV). Firstly, PhaGCN2 (v2.1) was used to classify the contigs (Jiang et al., 2023), while the unclassified contigs were annotated using geNomad (v1.8.0) (Camargo et al., 2023). The lifestyle of each phage was predicted by PhaTYP (v0.3.0) (Shang et al., 2023). The host of the phage was further predicted using iPHoP (v1.3.3) (Arumugam et al., 2023), and the candidates with highest-score were selected as potential hosts. AMGs in viral contigs were analyzed using VIBRANT (v1.2.1) (Kieft et al., 2020). All the above tools used the default settings. The flow of viral community analysis in this study was shown in Fig. S1.

2.7. Statistical analysis and visualization

Statistical analysis and visualization were performed using R (v4.3.0), Origin 2024, OmicStudio (https://www.omicstudio.cn/tool) and Chiplot (https://www.chiplot.online/). Significant differences between groups were assessed using one-way analysis of variance (ANOVA) followed by t-tests. Post hoc tests were conducted using Tukey's multiple comparisons.

3. Results

3.1. Virus identification and diversity analysis of fermented grains at different fermentation time

Of the 15 samples, 13 had received metagenomic sequencing, generating approximately 10 GB of raw data per sample. The other two were not sequenced due to low DNA content. The quality of the metagenomic sequencing is presented in Table S1. After quality filtering, the proportion of clean reads was 52.31 $\pm$ 12.41 % (Table S1). The average sequencing depth was approximately 50.9 million paired-end reads per sample (interquartile range: 44.4–64.4 million) (Table S1). Virus identification for the metagenome assembly contigs of the 13 samples was conducted using four tools (Fig. 2A and B). For subsequent analysis, we selected a combined set of 53,269 viral contigs recognized by all four tools as the candidate viral contigs. We used CheckV to perform quality control on the identified viral contigs, removing those with high host contamination and assessing viral contig integrity and quality. As a result, 6822 contigs were remained, including 60 high-quality, 249 medium-quality, and 70 complete circular phage contigs (Fig. 2B; Table S2). Furthermore, all contigs harboring at least one viral gene were clustered at a 95 % nucleotide identity threshold, yielding a total of 4105 viral operational taxonomic units (vOTUs) from the metagenomic data.

Diversity analysis of vOTUs obtained using metagenomics sequencing methods was conducted over time, and the results are presented in Fig. 2C and D. The Shannon diversity index of viral vOTUs varied across fermentation time points. A significant decrease was observed on day 5, followed by a gradual increase. The highest diversity was recorded on day 20, although no statistically significant differences were found among days 0, 10, 20, and 30 ($P > 0.05$). Non-metric multidimensional scaling (NMDS) analysis based on Bray-Curtis dissimilarity revealed distinct viral community structures at different fermentation stages (Stress = 0.0209). ANOSIM results showed significant differences among time points ($R = 0.645$, $P = 0.001$), indicating a temporal succession of viral community during fermentation.

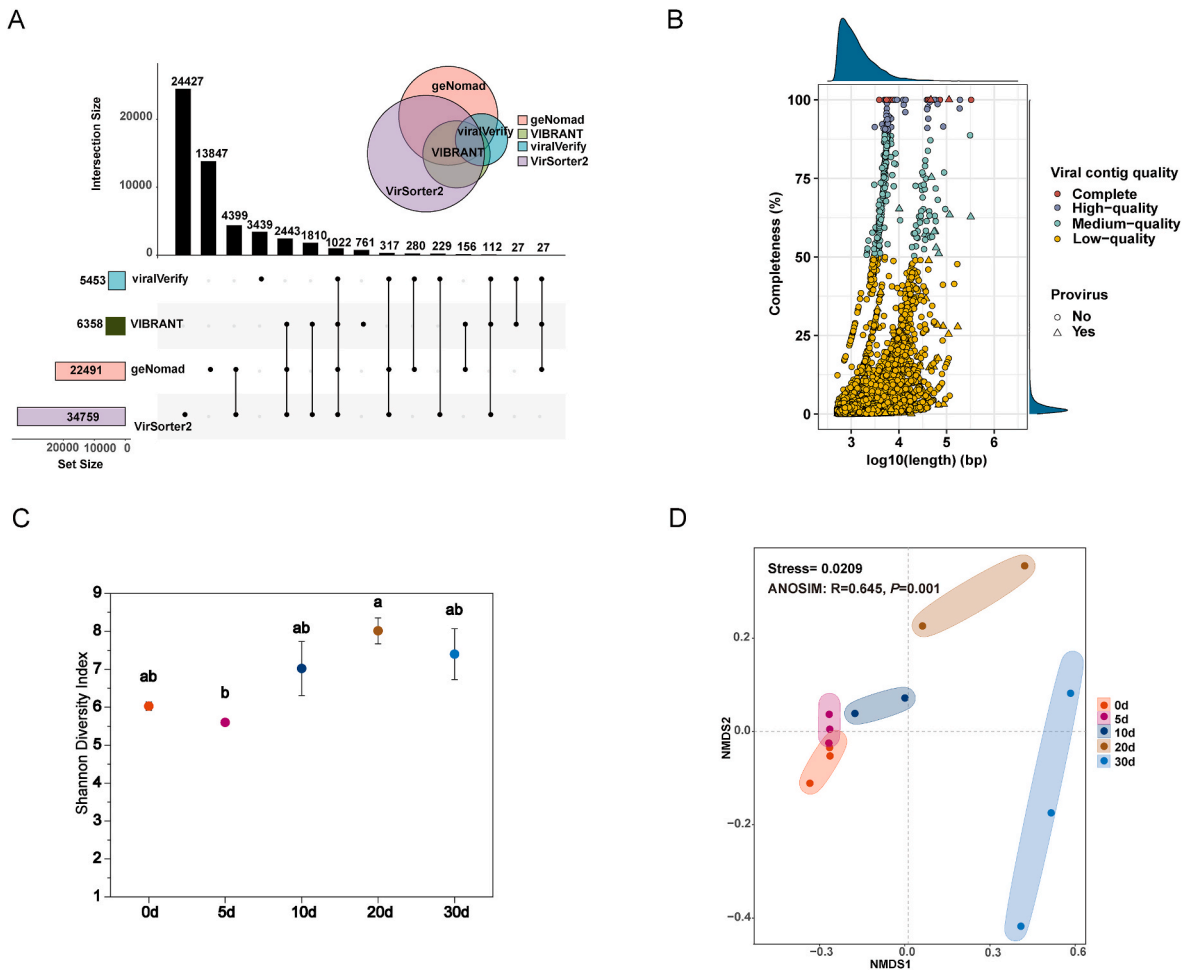

Fig. 2. Virus identification and diversity analysis. (A) Number of viruses identified using different tools. (B) Quantity and quality of viral contigs. (C) vOTU Shannon index across different fermentation times (mean ± standard deviation). Identical letters represent no significant differences, while different letters indicate significant differences ($P < 0.05$). (D) NMDS analysis of vOTU at different fermentation times.

3.2. The microbial and viral dynamics across five stages of the Baijiu fermentation

3.2.1. Microbial succession changes in the Baijiu fermentation

We investigated the relative abundances of bacteria, eukaryota, archaea, and viruses in samples collected at different fermentation time points (Fig. 3A). Overall, bacteria and eukaryotes exhibited high abundance throughout the fermentation process. Notably, bacteria were predominant on day 30, accounting for 73.08 ± 9.41 %, while eukaryotes were more abundant during the early fermentation stage (0–5 days), with a relative abundance of 33.34 ± 1.16 %. Although viruses were present at relatively low levels, they were consistently detected across all stages. Their abundance was higher during the first 20 days (4.05 ± 0.80 %) and dropped to the lowest level by day 30 (1.39 ± 0.91 %). Archaea were initially detected in trace amounts during the first 5 days (0.17 ± 0.08 %) but showed a marked increase after 10 days, reaching 1.43 ± 0.92 %.

The alpha diversity of microorganisms increased gradually during the early fermentation stage, indicating a rise in both species richness and evenness. However, a marked shift in community composition was observed at day 30 (Fig. S2A). Beta diversity analysis revealed clear clustering of samples by fermentation time point, with samples from day 30 showing a distinct separation from other time points, suggesting a substantial change in microbial community structure at this stage (Fig. S2B). We further examined the relative abundance of major microbial genera across different fermentation stages, identifying a total of 4352 genera (Table S3). The microbial community exhibited a dynamic succession pattern throughout the fermentation process. During the 0–10 day period, dominant genera included *Aspergillus* (10.00 ± 1.10 %), *Lichtheimia* (4.01 ± 0.96 %), *Paecilomyces* (3.75 ± 0.49 %), and *Weissella* (2.99 ± 1.46 %). At day 20, notable differences were observed between the two replicate samples, which may be attributed to variations in sampling depth. By day 30, *Acetilactobacillus* had become the dominant genus, accounting for 45.05 ± 25.73 % of the community (Fig. 3B).

3.2.2. Viral succession changes in the Baijiu fermentation

Furthermore, we analyzed the relative abundance of different virus types—ssRNA, dsRNA, ssDNA, and dsDNA—in each sample (Fig. 3C). The results revealed that ssRNA viruses (44.82 ± 6.56 %) and ssDNA viruses (27.68 ± 12.36 %) were more abundant during the early fermentation stage (0–5 days). In contrast, dsDNA viruses became dominant during the middle to late stages (10–30 days), with an average relative abundance of 70.59 ± 11.91 %. dsRNA viruses were consistently present at low levels throughout the process, with their highest abundance observed on day 30 (1.27 ± 0.50 %). These findings indicate that the viral community structure underwent significant shifts over the course of fermentation.

In the metagenomic data, 1381 of the 4105 vOTUs (33.64 %) were annotated at the family level, representing a total of 101 viral families (Fig. 3D; Table S4). We investigated changes of the top 10 dominant virus families during fermentation. Metaviridae primarily infect

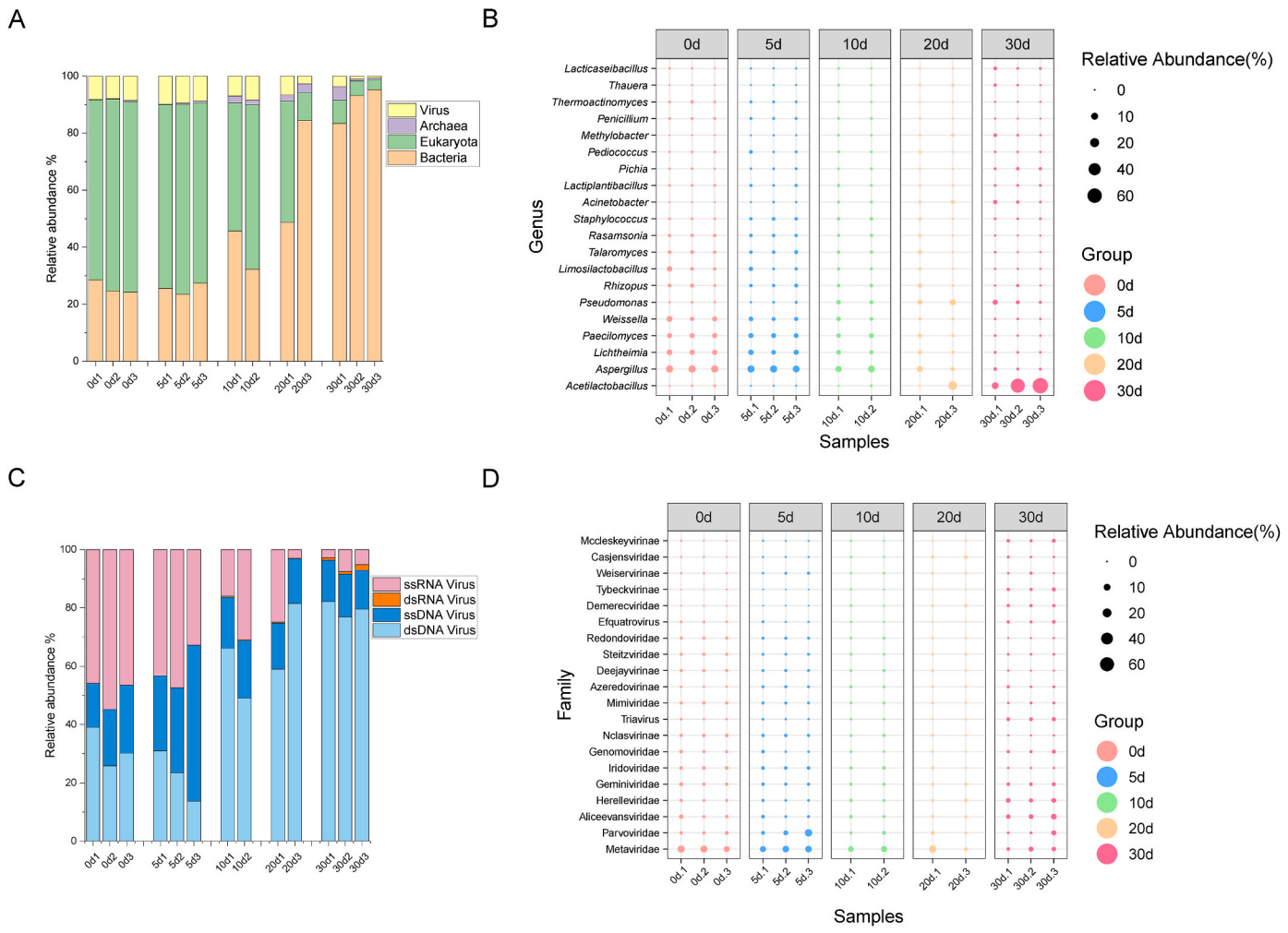

Fig. 3. Succession of microbial and viral communities during the fermentation of strong-flavor Baijiu. (A) Stacked bar plot showing the relative abundance of organisms at the domain level (Bacteria, Eukaryota, Archaea, and Virus) across different fermentation timepoints. (B) Bubble plot displaying the relative abundance of representative microbial genera in each sample; dot size corresponds to abundance, and color represents fermentation stage. (C) Relative abundance of virus types classified by genome structure (ssRNA, dsRNA, ssDNA, dsDNA) across samples. (D) Bubble plot illustrating the relative abundance of viral families at different time points, with dot size indicating abundance and color indicating group.

eukaryotes, including fungi, plants, and animals (Koonin et al., 2021). The abundance of the Metaviridae family gradually decreased during the fermentation process, from 9.83 ± 1.94 % at day 0, reaching its lowest point (2.89 ± 1.16 %) at day 30. The Parvoviridae family comprises small, single-stranded DNA viruses that predominantly infect mammals, birds, and insects (Cotmore et al., 2019). These viruses maintained stable persistence throughout the fermentation processes without showing significant variations (2.40 ± 1.81 %). Aliceevansviridae (dsDNA phages infecting Gram-negatives) showed no change in the first 20 fermentation days (1.14 ± 0.68 %) but rose substantially (4.27 ± 1.25 %) by day 30 (International Committee on Taxonomy of, 2023).

Herelleviridae is a phage family that infects bacteria, particularly those associated with environmental bacterial populations, and is classified under dsDNA virus (Barylski et al., 2020). In this study, the relative abundance of Herelleviridae did not show significant changes over the first 20 d (0.79 ± 0.28 %). However, at day 30, there was a significant increase, with a relative abundance of 3.23 ± 0.69 %. This may have been related to an increase in the host population of Herelleviridae on day 30.

Geminiviridae is a family of single-stranded circular DNA plant viruses (Fiallo-Olivé et al., 2021), which did not show significant changes during the fermentation process (1.30 ± 0.63 %). The Iridoviridae is a viral family that primarily infects invertebrates and is widely distributed in both aquatic and terrestrial environments. This family includes various viruses with dsDNA genomes (Khotimchenko & Shchelkanov, 2024). The abundance of Iridoviridae was higher at the beginning of fermentation (1.49 ± 0.06 % on day 0) and gradually decreased over time. Particularly after day 10 (1.25 ± 0.35 %), the abundance dropped quickly, maintaining low levels on day 30 (0.19 ± 0.02 %). This suggests that the activity of Iridoviridae gradually weakened during the fermentation process, which may be related to changes in the fermentation environment and microbial community succession.

Genomoviridae is a relatively new viral family comprising ssDNA viruses. This family typically infects plants, fungi, or bacteria, and is highly adaptable and capable of surviving in various ecosystems (Malathi and Renuka Devi, 2019). In the present study, the Genomoviridae abundance remained stable over time (0.75 ± 0.55 %). Nclasvirinae is a viral subfamily within the class Caudoviricetes, primarily comprising phages that infect bacteria (Zhu et al., 2022). We observed that the abundance of this viral subfamily gradually declined during the fermentation process, decreasing from 1.19 ± 0.08 % at the initial stage to 0.12 ± 0.01 % by day 30. Triavirus is a family of dsDNA phages that primarily infect Gram-positive bacteria, particularly *Staphylococcus aureus* (Ene et al., 2021). During the first 20 days of fermentation, no significant changes in abundance were observed (0.26 ± 0.08 %), but a sharp increase occurred by day 30 (1.86 ± 0.35 %), likely associated

with the rise in host population. Mimiviridae, known as "giant viruses," defy conventional viral definitions due to their exceptional particle size and genomic complexity. They possess linear dsDNA genomes ranging from approximately 617.4 kb to 1992.0 kb, encoding numerous genes, and exhibit a broad host range encompassing various eukaryotic organisms (Claverie and Abergel, 2018). During fermentation, the relative abundance of this viral family gradually declined, decreasing from an initial 0.89 ± 0.01 % to 0.23 ± 0.17 % by the end of the process. These data suggest temporal shifts in viral composition, which could be indicative of viral succession or varying environmental conditions influencing viral dynamics. Further investigations are required to understand the underlying ecological mechanisms driving these changes.

3.3. Changes in physicochemical parameters during Baijiu fermentation

To characterize the fermentation environment, we monitored key physicochemical indicators, including ethanol, organic acids (acetate and lactic acid), pH, temperature, glucose, and moisture across five time points (day 0, 5, 10, 20, and 30) (Fig. S3A–F). Ethanol concentration (Fig. S3A) rose sharply, especially between day 5 and day 10 (from 11.32 ± 2.18 to 21.50 ± 3.27 g/kg), and reached a maximum of 37.85 ± 5.44 g/kg by day 30, consistent with active saccharification and ethanol fermentation. Acetate levels (Fig. S3B) showed a progressive increase throughout the fermentation, from 1.71 ± 0.22 g/kg to day 0–6.53 ± 1.10 g/kg at day 30, indicating the accumulation of volatile acids over time. Lactic acid (Fig. S3C) showed a continuous upward trend from 24.82 ± 1.53 g/kg to 32.74 ± 2.67 g/kg over the fermentation period, highlighting the activity of lactic acid bacteria and their contribution to acidity. The pH (Fig. S3D) remained relatively stable (~3.55) during the first 20 days, but decreased slightly to 3.46 ± 0.12 by day 30, possibly due to organic acid accumulation in the late stage. Temperature (Fig. S3E) followed a typical fermentation profile, increasing from ~23.01 °C to a peak of 28.22 ± 3.53 °C at day 20, then declining to 25.15 ± 2.75 °C by day 30, likely reflecting microbial metabolic activity and heat generation. Glucose content (Fig. S3F) increased rapidly during the initial stages, peaking at day 10 (20.17 ± 4.64 g/kg), but subsequently declined to 4.25 ± 2.12 g/kg by day 30, indicating its consumption by fermentative microbes. Moisture content (Fig. 1G) exhibited a biphasic trend during the 30-day fermentation process. Initially, the moisture level slightly decreased from 59.02 ± 1.52 % at day 0 to a minimum of 57.53 ± 0.84 % at day 5, possibly due to early-stage evaporation or absorption by fermenting substrates. From

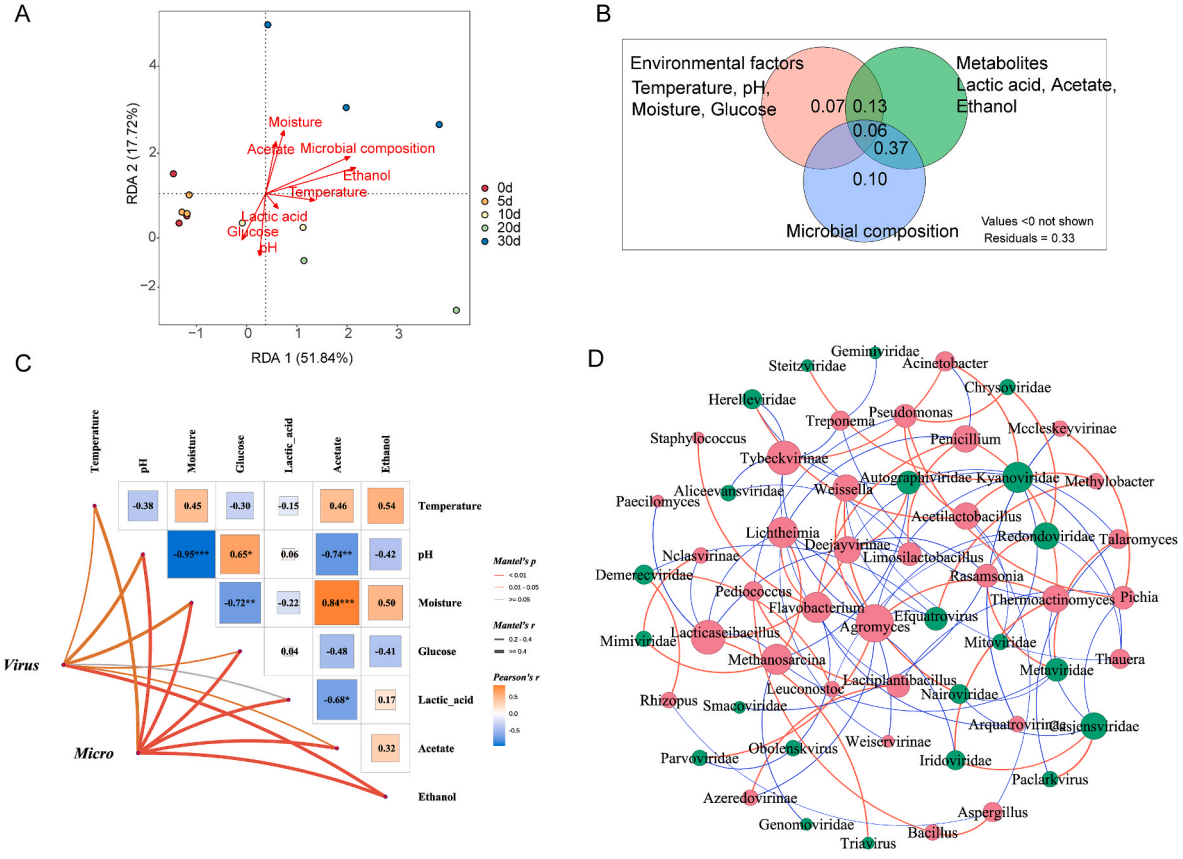

Fig. 4. Multivariate analyses linking viral communities with abiotic and biotic factors. (A) Redundancy analysis (RDA) plots shows the weights and orientation of viral communities and environmental parameters in fermentation process. To represent overall microbiome variation, the first axis (NMDS1) from NMDS based on Bray–Curtis distances was used as a proxy variable for microbial community composition. (B) Variation partitioning analysis (VPA) shows the contribution of environmental factors, metabolites and microbial community to the viral communities. (C) Correlation network and matrix analysis among environmental factors, metabolic products, and high-abundance viral families (Virus) and microbial genera (Micro). Upper triangular matrix displays Pearson's correlation coefficients (r) among physicochemical parameters. Blue and orange colors represent negative and positive correlations, respectively; asterisks denote significance levels (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). The chord diagram below represents Mantel test results between Virus/Micro and each environmental factor or metabolite. Edge color indicates p-value significance level, while edge thickness represents Mantel's r (correlation strength). (D) SparCC-based correlation network between microbial genera and viral families. The network displays significant correlations ($|\rho| > 0.3$, $P < 0.05$) between microbial genera (pink nodes) and viral families (green nodes) based on compositional data analysis using the SparCC algorithm. Node size represents the degree (number of connections), reflecting the centrality of each taxon in the network. Edges indicate significant correlations: red lines denote positive associations, and blue lines represent negative associations. The network layout was optimized to highlight microbial–viral interaction patterns, suggesting potential ecological relationships such as co-occurrence or mutual exclusion in the fermentation ecosystem.

day 10 onwards, moisture remained relatively stable (~58 %) until day 20, followed by a marked increase to 64.75 ± 5.61 % at day 30. This late-stage rise may result from the release of water due to microbial metabolism and substrate degradation. The increasing variability at day 30 also suggests heterogeneous water distribution in the fermented grain matrix during prolonged fermentation. Overall, these dynamic changes illustrate a typical fermentation trajectory, with increasing ethanol and organic acid levels, decreasing glucose availability, and stable-to-declining pH, all contributing to the shaping of the microbial and viral community structure.

3.4. The relationships among the detected virus, bacteria, fungi and environmental factors during the Baijiu fermentation

Redundancy analysis (RDA) with metagenome indicated that RDA1 and RDA2 explained 51.84 % and 17.72 % of the total variation in the viral community structure, respectively, accounting for nearly 70 % of the overall variance (Fig. 4A). To represent bacterial and fungal variation, the first axis (NMDS1) from NMDS based on Bray-Curtis distances was used as a proxy variable for microbial community composition (Fig. S2). Among the explanatory variables, Bacterial and fungal composition (Microbe_NMDS1) exhibited a significant contribution in the RDA, indicating that microbial community succession may serve as a key driver of viral community dynamics. Notably, Microbe_NMDS1 showed a similar ordination direction to ethanol, suggesting that these variables may jointly influence the co-succession patterns of viral communities. Variation partitioning analysis (VPA) revealed that the combined effect of environmental factors, metabolic products, and microbial community explained 73 % of the total variation in viral community structure, with the largest shared fraction (37 %) attributed to the interaction among all three groups (Fig. 4B). Notably, metabolic products independently contributed 13 % of the variation, indicating their distinct influence on virus. The remaining 27 % of unexplained variation (residuals = 0.33) likely reflects the influence of unmeasured factors and inherent ecological complexity.

Mantel test results indicated that the high-abundance viral family were significantly correlated with factors such as pH, glucose, and acetate (Fig. 4C), suggesting that viral community structure may be regulated by these environmental variables. In contrast, high-abundance microbial genera (bacterial and fungal composition) exhibited significant correlations with a broader range of physicochemical parameters, particularly acetate, moisture, and ethanol, indicating that microbial communities may play a more complex role in the dynamics of metabolic product transformation.

Results from the co-occurrence network revealed a total of 84 significant virus–microbe associations ($|r| > 0.03$, $P < 0.01$), comprising both positive (n = 42) and negative (n = 42) correlations (Fig. 4D). Among the most notable interactions, Casjensviridae displayed multiple associations, including a strong positive link with Paclarkvirus (r = 0.1087) and a negative correlation with *Pichia* (r = −0.0799), suggesting its potential involvement in virus–virus and virus–fungi interactions. *Agromyces* emerged as a microbial hub, showing both positive (e.g., with *Weissella*, r = 0.0731) and negative (e.g., with Redondoviridae, r = −0.0839) connections, indicating its possible dual roles in shaping viral community dynamics. Notably, Efquatrovirus exhibited a negative association with *Weissella* (r = −0.0984), while Metaviridae was positively correlated with *Talaromyces* (r = 0.0672) but negatively associated with Nairoviridae (r = −0.1056), highlighting complex interactions across different kingdoms. Several taxa such as *Lacticaseibacillus* and Tybeckvirinae were involved in both mutualistic and antagonistic relationships, reflecting a highly dynamic and modular network structure.

To identify strong associations between viruses and flavor compounds, we performed Spearman correlation analysis (Fig. S4). As shown in the network, multiple viral taxa—such as Mimiviridae, Fiersviridae, and Gclasvirinae—exhibited significant correlations with key flavor compounds, including ethyl acetate, pentanoic acid ethyl ester, butanoic acid ethyl ester, and 1-hexanol. Among these, 81.3 % were positively correlated and 18.7 % negatively correlated. These findings suggest that viruses may indirectly influence the production of flavor compounds by modulating the community structure or metabolic activity of their microbial hosts (e.g., aroma-producing yeasts or bacteria). This analysis provides preliminary evidence for a potential link between viral dynamics and flavor metabolism, expanding our understanding of the role of viral ecology in shaping the flavor profile of fermented systems.

3.5. The life style and host prediction of the detected phages

We investigated the changes in the relative abundance of phage lifestyles during the fermentation process (Fig. 5A). In the early stage (0–10 days), virulent phages were predominant, accounting for over 70 % of the total phage population. However, as fermentation progressed—particularly in samples from days 20 and 30—the proportion of temperate phages increased markedly, reaching or exceeding 50 % in some samples on day 30. This trend suggests that increased host stress and greater environmental stability in the later fermentation stages may favor the persistence or induction of temperate phages.

In the metagenomic data, most host-associated phages (e.g., those targeting *Limosilactobacillus*, *Lactobacillus*, *Weissella*, and *Staphylococcus*) were predominantly temperate. In contrast, phages associated with a few hosts such as *Pseudomonas*, *Acinetobacter*, and *Thermoactinomyces* showed nearly equal or slightly higher numbers of virulent types (Fig. 5B). These results demonstrated that in the fermentation system, phages infecting Gram-positive lactic acid bacteria were primarily temperate, which may contribute to host stability, whereas Gram-negative bacteria were more susceptible to attacks by lytic phages.

To further investigate the relationship between phages and their hosts, we calculated the linear correlation coefficients between the relative abundances of phages and their corresponding hosts during fermentation (Fig. 5C). Among the twelve phages with the most frequently identified hosts, all but *Lacticaseibacillus* phage exhibited a strong positive linear correlation with their hosts. This suggests a concomitant relationship between these phages and their hosts during Baijiu fermentation, indicating that an increase in host abundance is associated with an increase in phage abundance. This relationship may be attributed to the fact that most of these phages are temperate, meaning that they do not lyse their hosts, consistent with the " piggyback-the-Winner " (PtW) hypothesis observed in ecological studies (Brown et al., 2022). The "PtW" strategy is a well-recognized concept wherein temperate phages preferentially infect and propagate within dominant microbial populations, significantly influencing the succession of microbial communities (Brown et al., 2022). During Baijiu fermentation, this mechanism may shape the predominance of specific microbial strains by enabling the propagation of phages within abundant host populations. By fostering the success of particular strains such as *Weissella*, *Lactiplantibacillus*, *Pseudomonas*, *Limosilactobacillus*, and *Staphylococcus*, this strategy likely plays a critical role in driving microbial succession during the fermentation process.

3.6. Functional gene annotation of virus

We annotated the functional genes of viruses and found that a substantial proportion remained uncharacterized (42.25 %). Among the annotated genes, those associated with viral infection and replication processes accounted for 15.23 %. In addition, a considerable number of genes were related to metabolic functions, including amino acid, carbohydrate, coenzyme, and lipid metabolism (Fig. S5). In this study, 162 AMGs were predicted during the fermentation of strong-flavor Baijiu (Fig. S6A). These AMGs were found to be extensively involved in the metabolism of carbohydrates, amino acids, cofactors, vitamins, and in other essential processes. The complete phage sequence (Kyanoviridae, virulent) was 320 kb in length and encoded 445 ORFs, of which 48 had

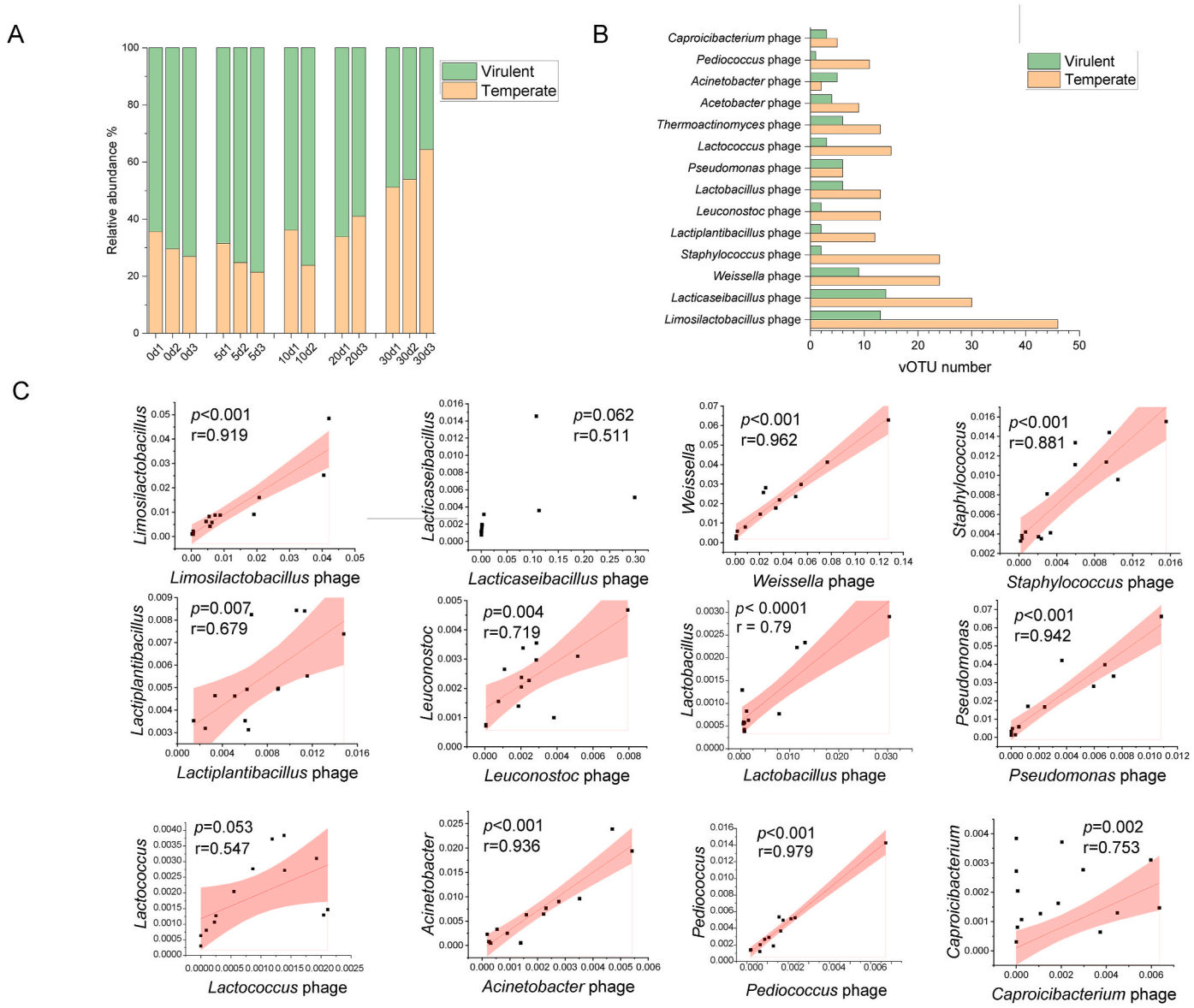

Fig. 5. Lifestyle distribution and host correlations of phages during fermentation. (A) Relative abundance of virulent and temperate phages across different fermentation time. (B) Number of virulent and temperate phage vOTUs associated with different bacterial genera. (C) Changes in relative abundance of phages and their corresponding hosts in metagenomics data. The "*p*" represents the significance level of the statistical tests, with $P < 0.05$ indicating statistically significant results. The "r" value indicates the Pearson correlation coefficient, which quantifies the strength and direction of the linear relationship between the relative abundances of phages and their corresponding hosts. A higher absolute value of "r" signifies a stronger correlation.

functional annotations, including two viral hallmark genes related to the major capsid protein Gp23 and HNH endonuclease (Fig. S6B). In addition, we identified nine AMGs in high-quality vOTU (Fig. S6C). Notably, we identified an AMG-encoded *DNMT1*, a DNA (cytosine-5)-methyltransferase that protects viruses from their host antiviral restriction-modification system (Hampton et al., 2020; Murphy et al., 2013), which is consistent with the high prevalence of this type of AMG found in marine and rumen viromes (Heyerhoff et al., 2022; Yan et al., 2023). Notably, these AMGs were identified with 100 % Phyre2 confidence scores (Fig. S6D), indicating their potential for expression.

To further investigate these (AMGs), we analyzed their associated metabolic pathways and predicted host taxa (Fig. 6). A total of 31 key AMGs were identified, encompassing pathways involved in amino acid metabolism (e.g., arginine and proline metabolism; cysteine and methionine metabolism), energy metabolism (e.g., sulfur metabolism; butyric acid metabolism), nucleotide metabolism (e.g., purine metabolism), and coenzyme metabolism (e.g., nicotinate and nicotinamide metabolism), among others. Several AMGs (e.g., DNMT1, nadE, pncB, folA) were predicted to be associated with multiple host genera, including *Lactobacillus*, *Pediococcus*, *Pseudomonas*, and *Thermoactinomyces*, suggesting that viruses may influence host metabolic networks or enhance host adaptability by encoding functional metabolic genes. Notably, certain metabolic pathways—such as nicotinate and nicotinamide metabolism and polyketide sugar unit biosynthesis—exhibited extensive connections between multiple AMGs and diverse host genera, indicating that viruses may participate in multi-directional metabolic interactions within the microbial ecosystem.

Furthermore, we explored the variation in the abundance of the detected AMGs across different samples, represented by the abundance of the contigs containing the genes (Fig. S7; Table S6). We observed substantial differences in AMG abundance between the early (first 5 d) and later samples. For instance, AMGs involved in amino acid metabolism, such as *aroD* (3-dehydroquinate dehydratase), *glsA* (glutaminase), and *thrB1* (homoserine kinase), as well as those related to

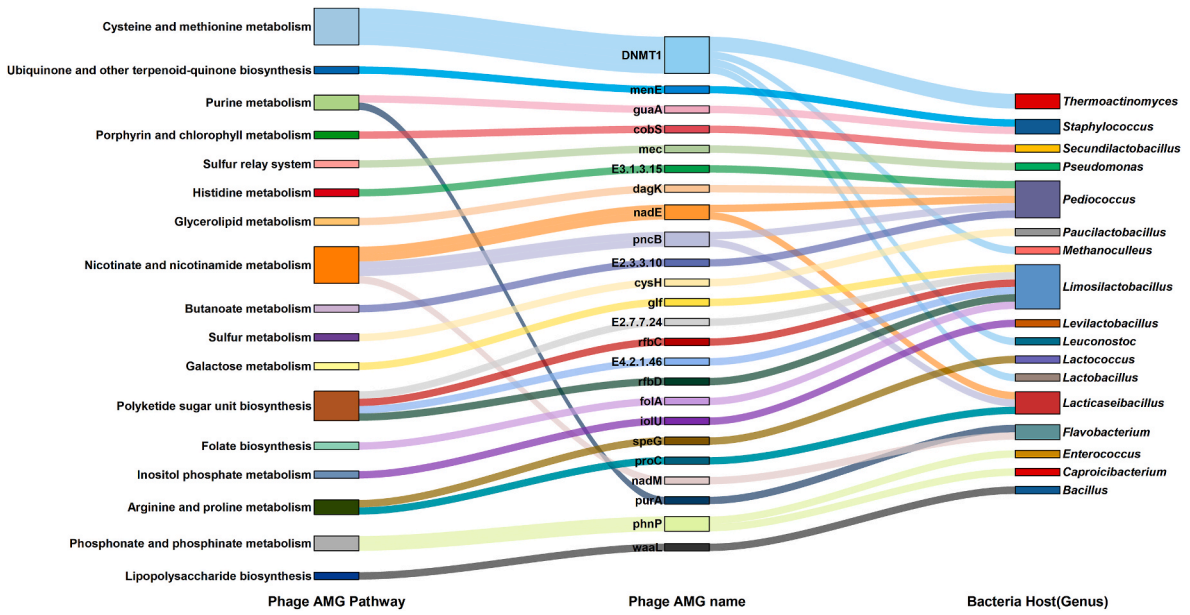

Fig. 6. Sankey diagram illustrating the distribution of phage AMGs across metabolic pathways, gene functions, and bacterial host genera. The diagram links phage AMGs (center) to their corresponding metabolic pathways (left) and bacterial host genera (right). The thickness of the connecting line indicates the number of AMG.

xenobiotics biodegradation and metabolism (*CYP53A1*, benzoate 4-monooxygenase), purine metabolism (*URE*, urease), energy metabolism (*metF*, methylenetetrahydrofolate reductase), and carbohydrate metabolism (*E3.5.1.41*, chitin deacetylase) showed higher abundance in the early samples but significantly decreased after 10 d. In contrast, AMGs related to carbohydrate metabolism such as *neuA* (N-acylneuraminate cytidylyltransferase), *UXS1* (UDP-glucuronate decarboxylase), K15024 (putative phosphotransacetylase), *DLAT* (pyruvate dehydrogenase E2 component), *gpmB* (probable phosphoglycerate mutase), and AMGs involved in metabolism of cofactors and vitamins such as *ubiG* (2-polyprenyl-6-hydroxyphenyl methylase) and *pncA* (nicotinamidase/pyrazinamidase) exhibited higher abundance in samples collected after 10 d. In samples collected after 30 d of fermentation, the number and abundance of AMGs significantly increased compared with those at other time points. These included genes such as *nadD* (nicotinate-nucleotide adenylyltransferase), *ubiG* (2-polyprenyl-6-hydroxyphenyl methylase), phoD (alkaline phosphatase D), *ACADS* (butyryl-CoA dehydrogenase), *ACSL* (long-chain acyl-CoA synthetase), *dapB* (4-hydroxy-tetrahydrodipicolinate reductase), *fabG* (3-oxoacyl-[acyl-carrier protein] reductase), *gdh* (glucose dehydrogenase), *neuA* (N-acylneuraminate cytidylyltransferase), and *waaL* (O-antigen ligase).

By annotating the metabolic pathway of the *mec* gene, we found that it encodes a key enzyme involved in the serine-to-cysteine biosynthesis pathway (Fig. S8A and B). To validate the functional activity of the identified Mec protein, we conducted heterologous expression and enzymatic activity assays. The results showed that the protein had a molecular weight of approximately 30 kDa, consistent with the theoretical prediction, and was capable of catalyzing cysteine synthesis at both pH 4.5 and pH 7.5 (Fig. S8C and D).

4. Discussion

To our knowledge, this study represents the first comprehensive investigation of viral community involved in the fermentation of strong-flavor Baijiu using metagenomic analyses. From 13 fermented grain samples across five stages (days 0, 5, 10, 20, and 30), 101 viral families were identified, with the community dominated by members of Metaviridae, Parvoviridae, Aliceevansviridae, Herelleviridae, Geminiviridae, Iridoviridae, and Genomoviridae. The temporal dynamics of viral communities were closely linked to microbial assembly and ethanol production during fermentation. Additionally, we observed a significant positive correlation between major phages and their hosts, suggesting synchronized abundance patterns during fermentation. Moreover, the viral community harbors a diverse array of AMGs related to amino acid metabolism and vitamin biosynthesis. These insights establish a foundation for future research on viral communities in the production of Baijiu and contribute to a deeper understanding of the ecological roles of viruses in spontaneously fermented foods.

Consistent with findings from pickle and cheese fermentation systems, most of the identified viral hosts in our study were lactic acid bacteria (Cantuti Gendre et al., 2025; Paillet et al., 2024). In addition, we also detected viruses associated with the *Staphylococcus* genus during the fermentation process, which may be attributed to the open fermentation conditions. However, *Staphylococcus* are not a dominant functional bacterium in Baijiu fermentation (Wu et al., 2021). While certain *Staphylococcus* strains are used as functional bacteria in fermented meat products (R. Organji et al., 2018), their role in Baijiu fermentation remains limited. Notably, we identified phages infecting caproic acid–producing bacteria (Fig. 5B), which sets our system apart from pickle and cheese fermentations. These bacteria are key anaerobes responsible for caproic acid production in the cellar mud of strong-flavor Baijiu, and both their abundance and metabolic activity are closely linked to the final liquor quality (Gao et al., 2022).

The lysogenic cycle of phages may play a critical role in shaping microbial communities in spontaneously fermented foods (Zhang et al., 2024). Our findings suggest that a lysogenic lifestyle is preferentially adopted by microbes and prophages in fermented grains. During Baijiu fermentation, microbial communities undergo dynamic changes that may be influenced by the interactions between temperate phages and their hosts. These phages can act as selective agents, driving competition and succession within microbial populations and consequently affecting the metabolic outcomes of fermentation. The observed positive correlation between phages and their hosts supports the dominance of key bacteria such as *Weissella*, *Lactiplantibacillus*, *Pseudomonas*, and *Limosilactobacillus*, which are essential for producing the desired fermentation metabolites (Hu et al., 2022). Thus exploring this intricate interplay further might enhance understanding of microbial community succession in Baijiu production. A limitation of this study is the inability to obtain species-level host information, which may hinder a comprehensive exploration of the virus's host spectrum. Future research should

focus on the isolation and screening of microbial viruses involved in Baijiu fermentation.

In this study, we observed the significant presence of AMGs related to amino acid, vitamin, and cofactor metabolism (Fig. S6A), suggesting a unique contribution of phages to the Baijiu fermentation process. We identified putative AMGs from metagenomic data and performed gene cluster analysis as well as protein tertiary structure prediction (Fig. S6C and D). This study also identified a viral AMG involved in amino acid metabolism and verified its activity via heterologous expression (Fig. S8). To strengthen these findings, future work should involve metatranscriptomics (Zhang et al., 2024) and metabolomics to clarify the functional roles of viral AMGs. More rigorous validation would require viral isolation, culture, and infection assays in culturable host systems to directly assess the auxiliary metabolic roles of viruses.

5. Conclusions

This study employed metagenomics approaches to characterize the viral community during the fermentation of strong-flavor Baijiu. A total of 101 viral families were identified, with lactic acid bacteria predicted as the primary phage hosts. Our findings suggested that the co-succession of microbial and viral communities was driven by the combined effects of environmental conditions and metabolic activity, with complex virus–microbe interactions reflecting the ecological intricacy of fermented systems. Many viral AMGs related to amino acid metabolism have potential effects on the host. This study provides new insights into the often-overlooked roles of viruses in Baijiu fermented grains, highlighting their importance in fermented grain microbiota.

Data statement

Data will be made available on request.

Author contributions

Huadong Zhang – Data curation, Methodology, Writing – original draft.

Hongxia Zhang – Methodology, Project administration, Resources, Writing – review & editing.

Hai Du – Resources, Project administration.

Yan Zhang – Methodology, Project administration, Writing – review & editing.

Menghui Zhang – Resources, Methodology, Project administration, Writing – review & editing.

Xiaowei Yu – Resources, Project administration.

Yan Xu – Methodology, Project administration, Resources, Writing – review & editing.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC) (grants 32201981; 32172176).

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crfs.2025.101189.

Data availability

Data will be made available on request.

References

Antipov, D., Raiko, M., Lapidus, A., Pevzner, P.A., 2020. Metaviral SPAdes: assembly of viruses from metagenomic data. Bioinformatics 36 (14), 4126–4129.

Arumugam, M., Roux, S., Camargo, A.P., Coutinho, F.H., Dabdoub, S.M., Dutilh, B.E., Tritt, A., 2023. iPHoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. PLoS Biol. 21 (4). https://doi.org/10.1371/journal.pbio.3002083.

Barylski, J., Kropinski, A.M., Alikhan, N.F., Adriaenssens, E.M., 2020. ICTV virus taxonomy profile: herelleviridae. J. Gen. Virol. 101 (5), 362–363. https://doi.org/10.1099/jgv.0.001354.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120.

Brown, T.L., Charity, O.J., Adriaenssens, E.M., 2022. Ecological and functional roles of bacteriophages in contrasting environments: marine, terrestrial and human gut. Curr. Opin. Microbiol. 70, 102229. https://doi.org/10.1016/j.mib.2022.102229.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12 (1), 59–60.

Cai, W., Wang, Y., Liu, Z., Liu, J., Hou, Q., Yang, X., Guo, Z., 2022. Depth-depended quality comparison of light-flavor fermented grains from two fermentation rounds. Food Res. Int. 159, 111587.

Callanan, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P., Hill, C., 2021. Biases in viral metagenomics-based detection, cataloguing and quantification of bacteriophage genomes in human faeces, a review. Microorganisms 9 (3), 524.

Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Kyrpides, N.C., 2023. Identification of mobile genetic elements with geNomad. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-01953-y.

Cantuti Gendre, J., Le Marrec, C., Chaillou, S., Omhover-Fougy, L., Landaud, S., Dugat-Bony, E., 2025. Exploring viral diversity in fermented vegetables through viral metagenomics. Food Microbiol. 128. https://doi.org/10.1016/j.fm.2025.104733.

Claverie, J.M., Abergel, C., 2018. Mimiviridae: an expanding family of highly diverse large dsDNA viruses infecting a wide phylogenetic range of aquatic eukaryotes. Viruses 10 (9). https://doi.org/10.3390/v10090506.

Cotmore, S.F., Agbandje-McKenna, M., Canuti, M., Chiorini, J.A., Eis-Hubinger, A.M., Hughes, J., Ictv Report, C., 2019. ICTV virus taxonomy profile: parvoviridae. J. Gen. Virol. 100 (3), 367–368. https://doi.org/10.1099/jgv.0.001212.

Du, H., Chen, B., Fu, W., Yang, F., Lv, X., Tan, Y., Xu, Y., 2022. Composition and function of viruses in sauce-flavor baijiu fermentation. Int. J. Food Microbiol. 387, 110055. https://doi.org/10.1016/j.ijfoodmicro.2022.110055.

Ene, A., Miller-Ensminger, T., Mores, C., Giannattasio-Ferraz, S., Wolfe, A., Abouelfetouh, A., Putonti, C., 2021. Examination of Staphylococcus aureus prophages circulating in Egypt. Viruses 13 (2). https://doi.org/10.3390/v13020337.

Fiallo-Olivé, E., Lett, J.M., Martin, D.P., Roumagnac, P., Varsani, A., Zerbini, F.M., Navas-Castillo, J., 2021. ICTV Virus Taxonomy Profile: geminiviridae 2021. J. Gen. Virol. 102 (12). https://doi.org/10.1099/jgv.0.001696.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28 (23), 3150–3152.

Gao, J., Qin, J., Ye, F., Ding, F., Liu, G., Li, A., Xu, Y., 2022. Constructing simplified microbial consortia to improve the key flavour compounds during strong aroma-type Baijiu fermentation. Int. J. Food Microbiol. 369. https://doi.org/10.1016/j.ijfoodmicro.2022.109594.

Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O., Roux, S., 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9 (1), 37. https://doi.org/10.1186/s40168-020-00990-y.

Hampton, H.G., Watson, B.N., Fineran, P.C., 2020. The arms race between bacteria and their phage foes. Nature 577 (7790), 327–336.

Heyerhoff, B., Engelen, B., Bunse, C., 2022. Auxiliary metabolic gene functions in pelagic and benthic viruses of the Baltic Sea. Front. Microbiol. 13, 863620.

Hu, X., Tian, R., Wang, K., Cao, Z., Yan, P., Li, F., He, P., 2021. The prokaryotic community, physicochemical properties and flavors dynamics and their correlations in fermented grains for Chinese strong-flavor Baijiu production. Food Res. Int. 148. https://doi.org/10.1016/j.foodres.2021.110626.

Hu, Y., Zhang, L., Wen, R., Chen, Q., Kong, B., 2022. Role of lactic acid bacteria in flavor development in traditional Chinese fermented foods: a review. Crit. Rev. Food Sci. Nutr. 62 (10), 2741–2755.

Huang, X., Li, R., Xu, J., Kang, J., Chen, X., Han, B., Xue, Y., 2025. Integrated multi-omics uncover viruses, active fermenting microbes and their metabolic profiles in the Daqu microbiome. Food Res. Int. https://doi.org/10.1016/j.foodres.2025.116061.

Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., Schuster, S.C., 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res. 21 (9), 1552–1560.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11, 1–11.

International Committee on Taxonomy of, V, 2023. ICTV virus taxonomy profile: aliceevansviridae. ICTV Taxonomy Release 2023 (March 2023 Release). N/a. https://ictv.global/taxonomy/taxondetails?taxnode_id=202309022.

Jiang, J.-Z., Yuan, W.-G., Shang, J., Shi, Y.-H., Yang, L.-L., Liu, M., Yuan, L.-H., 2023. Virus classification for viral genomic fragments using PhaGCN2. Briefings Bioinf. 24 (1). https://doi.org/10.1093/bib/bbac505.

Kang, J., Chen, X., Han, B.Z., Xue, Y., 2022. Insights into the bacterial, fungal, and phage communities and volatile profiles in different types of Daqu. Food Res. Int. 158, 111488. https://doi.org/10.1016/j.foodres.2022.111488.

Khotimchenko, Y.S., Shchelkanov, M.Y., 2024. Viruses of the Ocean: on the shores of the aqua Incognita. Horizons of taxonomic diversity. Russ. J. Mar. Biol. 50 (1), 1–24. https://doi.org/10.1134/S106307402401005X.

Kieft, K., Zhou, Z., Anantharaman, K., 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome 8 (1), 90. https://doi.org/10.1186/s40168-020-00867-0.

Koonin, E.V., Dolja, V.V., Krupovic, M., Kuhn, J.H., 2021. Viruses defined by the position of the virosphere within the replicator space. Microbiology and molecular biology reviews 85 (4). https://doi.org/10.1128/mmbr.00193-20.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31 (10), 1674–1676.

Liu, S., Moon, C.D., Zheng, N., Huws, S., Zhao, S., Wang, J., 2022. Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. Microbiome 10 (1), 76.

Luo, L.-J., Song, L., Han, Y., Zhen, P., Han, D.-Y., Zhao, X., Han, P.-J., 2023. Microbial communities and their correlation with flavor compound formation during the mechanized production of light-flavor Baijiu. Food Res. Int. 172, 113139.

Ma, J., Qian, C., Hu, Q., Zhang, J., Gu, G., Liang, X., Zhang, L., 2024. The bacteriome-coupled phage communities continuously contract and shift to orchestrate the traditional rice vinegar fermentation. Food Res. Int. 184. https://doi.org/10.1016/j.foodres.2024.114244.

Malathi, V.G., Renuka Devi, P., 2019. ssDNA viruses: key players in global virome. VirusDisease 30 (1), 3–12. https://doi.org/10.1007/s13337-019-00519-4.

Murphy, J., Mahony, J., Ainsworth, S., Nauta, A., van Sinderen, D., 2013. Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. Appl. Environ. Microbiol. 79 (24), 7547–7555.

Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S., Kyrpides, N.C., 2020. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat. Biotechnol. 39 (5), 578–585. https://doi.org/10.1038/s41587-020-00774-7.

Organji, R.S., Abulreesh, H.H., Elbanna, K., Osman, G.E.H., Almalki, M.H.K., 2018. Diversity and characterization of staphylococcus Spp. In food and dairy products: a foodstuff safety assessment. J. Microbiol. Biotechnol. Food Sci. 7 (6), 586–593. https://doi.org/10.15414/jmbfs.2018.7.6.586-593.

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Kyrpides, N.C., 2016. Uncovering Earth's virome. Nature 536 (7617), 425–430. https://doi.org/10.1038/nature19094.

Paillet, T., Lamy-Besnier, Q., Figueroa, C., Petit, M.A., Dugat-Bony, E., 2024. Dynamics of the viral community on the surface of a French smear-ripened cheese during maturation and persistence across production years. mSystems 9 (7). https://doi.org/10.1128/msystems.00201-24.

Santos-Medellin, C., Zinke, L.A., Ter Horst, A.M., Gelardi, D.L., Parikh, S.J., Emerson, J. B., 2021. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. ISME J. 15 (7), 1956–1970. https://doi.org/10.1038/s41396-021-00897-y.

Shang, J., Tang, X., Sun, Y., 2023. PhaTYP: predicting the lifestyle for bacteriophages using BERT. Brief Bioinform 24 (1). https://doi.org/10.1093/bib/bbac487.

Sommers, P., Chatterjee, A., Varsani, A., Trubl, G., 2021. Integrating viral metagenomics into an ecological framework. Annu. Rev. Virol. 8 (1), 133–158.

Song, Z., Du, H., Zhang, M., Nie, Y., Xu, Y., 2019. Schizosaccharomyces pombe can reduce acetic acid produced by baijiu spontaneous fermentation Microbiota. Microorganisms 7 (12). https://doi.org/10.3390/microorganisms7120606.

Steinegger, M., Söding, J., 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. 35 (11), 1026–1028. https://doi.org/10.1038/nbt.3988.

Tan, Y., Zhong, H., Zhao, D., Du, H., Xu, Y., 2019. Succession rate of microbial community causes flavor difference in strong-aroma Baijiu making process. Int. J. Food Microbiol. 311, 108350. https://doi.org/10.1016/j.ijfoodmicro.2019.108350.

Trubl, G., Hyman, P., Roux, S., Abedon, S.T., 2020. Coming-of-age characterization of soil viruses: a user's guide to virus isolation, detection within metagenomes, and viromics. Soil Syst. 4 (2), 23.

Wang, S., Xiong, W., Wang, Y., Nie, Y., Wu, Q., Xu, Y., Geisen, S., 2020. Temperature-Induced annual variation in microbial community changes and resulting metabolome shifts in a controlled fermentation System. mSystems 5 (4). https://doi.org/10.1128/mSystems.00555-20. Article e00555-20.

Wang, G., Li, S., Yan, Q., Guo, R., Zhang, Y., Chen, F., Ma, X., 2023. Optimization and evaluation of viral metagenomic amplification and sequencing procedures toward a genome-level resolution of the human fecal DNA virome. J. Adv. Res. 48, 75–86.

Wei, J.L., Lu, J., Nie, Y., Li, C.W., Du, H., Xu, Y., 2023. Amino acids drive the deterministic assembly process of fungal community and affect the flavor metabolites in baijiu fermentation. Microbiol. Spectr. https://doi.org/10.1128/spectrum.02640-22.

Wu, Q., Zhu, Y., Fang, C., Wijffels, R.H., Xu, Y., 2021. Can we control microbiota in spontaneous food fermentation? – chinese liquor as a case example. Trends Food Sci. Technol. 110, 321–331. https://doi.org/10.1016/j.tifs.2021.02.011.

Yan, M., Pratama, A.A., Somasundaram, S., Li, Z., Jiang, Y., Sullivan, M.B., Yu, Z., 2023. Interrogating the viral dark matter of the rumen ecosystem with a global virome database. Nat. Commun. 14 (1). https://doi.org/10.1038/s41467-023-41075-2.

You, L., Zhao, D., Zhou, R., Tan, Y., Wang, T., Zheng, J., 2021. Distribution and function of dominant yeast species in the fermentation of strong-flavor baijiu. World J. Microbiol. Biotechnol. 37, 1–12.

Yu, Z., Ma, Y., Guan, Y., Zhu, Y., Wang, K., Wang, Y., Yu, Y., 2022. Metagenomics of virus diversities in solid-state brewing process of traditional Chinese vinegar. Foods 11 (20). https://doi.org/10.3390/foods11203296.

Yuan, S.K., Du, H., Zhao, D., Qiao, Z.W., Zheng, J., Yu, X.W., Xu, Y., 2023. Stochastic processes drive the assembly and metabolite profiles of keystone taxa during Chinese strong-flavor baijiu fermentation. Microbiol. Spectr. https://doi.org/10.1128/spectrum.05103-22.

Zhang, H., Zhang, H., Du, H., Yu, X., Xu, Y., 2024. The insights into the phage communities of fermented foods in the age of viral metagenomics. Crit. Rev. Food Sci. Nutr. 1–13. https://doi.org/10.1080/10408398.2023.2299323.

Zhang, J., Tang, A., Jin, T., Sun, D., Guo, F., Lei, H., Li, B., 2024. A panoramic view of the virosphere in three wastewater treatment plants by integrating viral-like particle-concentrated and traditional non-concentrated metagenomic approaches. iMeta. https://doi.org/10.1002/imt2.188.

Zhang, H., Zhang, H., Du, H., Zhang, Y., Zhang, M., Yu, X., Xu, Y., 2025. Unraveling the multiple interactions between phages, microbes and flavor in the fermentation of strong-flavor Baijiu. Bioresources and Bioprocessing 12 (1). https://doi.org/10.1186/s40643-025-00852-1.

Zhu, Y., Shang, J., Peng, C., Sun, Y., 2022. Phage family classification under caudoviricetes: a review of current tools using the latest ICTV classification framework. Front. Microbiol. 13. https://doi.org/10.3389/fmicb.2022.1032186.