



OPEN ACCESS

EDITED BY

Haifeng Zhao,
South China University of Technology, China

REVIEWED BY

Fang Fang,
Jiangnan University, China
Chongde Wu,
Sichuan University, China
Zhen-Lin Xu,
South China Agricultural University, China
Kai Zhou,
Jiujiang University, China

*CORRESPONDENCE

Guiliang Tan
✉ joe88tan@126.com
Min Hu
✉ humin@cczu.edu.cn

RECEIVED 26 October 2023

ACCEPTED 05 December 2023

PUBLISHED 22 December 2023

CITATION

Tan G, Qi S, Wang Y, Li X, Li X, Li M, Li L, Zhao L and Hu M (2023) Uncovering differences in the composition and function of phage communities and phage-bacterium interactions in raw soy sauce. *Front. Microbiol.* 14:1328158. doi: 10.3389/fmicb.2023.1328158

COPYRIGHT

© 2023 Tan, Qi, Wang, Li, Li, Li, Zhao and Hu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Uncovering differences in the composition and function of phage communities and phage-bacterium interactions in raw soy sauce

Guiliang Tan^{1*}, Shaohan Qi², Yi Wang¹, Xueyan Li¹, Xiangli Li³, Mei Li¹, Lin Li¹, Lichao Zhao² and Min Hu^{4*}

¹School of Material Science and Food Engineering, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan, China, ²College of Food Science, South China Agricultural University, Guangzhou, Guangdong, China, ³School of Health Industry, Zhongshan Torch Polytechnic, Zhongshan, China, ⁴School of Environmental Science and Engineering, Changzhou University, Changzhou, China

Introduction: Although the composition and succession of microbial communities in soy sauce fermentation have been well-characterized, the understanding of phage communities in soy sauce remains limited.

Methods: This study determined the diversity, taxonomic composition, and predicted function of phage communities and the phage-host interactions in two types of raw soy sauce (Cantonese-type fermentation, NJ; Japanese-type fermentation, PJ) using shotgun metagenomics.

Results and discussion: These two raw soy sauces showed differences in phage composition (121 viral operational taxonomic units (vOTUs) in NJ and 387 vOTUs in PJ), with a higher abundance of the family *Siphoviridae* (58.50%) in the NJ phage community and a higher abundance of *Myoviridae* (33.01%) in PJ. Auxiliary metabolic functional annotation analyses showed that phages in the raw soy sauces mostly encoded genes with unknown functions (accounting for 66.33% of COG profiles), but the NJ sample contained genes mostly annotated to conventional functions related to carbohydrate metabolism (0.74%) and lipid metabolism (0.84%), while the PJ sample presented a higher level of amino acid metabolism functions (0.12%). Thirty auxiliary metabolism genes (AMGs) were identified in phage genomes, which were associated with carbohydrate utilization, cysteine and methionine metabolism, and aspartic acid biosynthesis for the host. To identify phage-host interactions, 30 host genomes (affiliated with 22 genera) were also recruited from the metagenomic dataset. The phage-host interaction analysis revealed a wide range of phage hosts, for which a total of 57 phage contigs were associated with 17 host genomes, with *Shewanella fodinae* and *Weissella cibaria* infected by the most phages. This study provides a comprehensive understanding of the phage community composition, auxiliary metabolic functions, and interactions with hosts in two different types of raw soy sauce.

KEYWORDS

raw soy sauce, phage diversity and functions, phage-host association, metagenome-assembled genome, auxiliary metabolic genes

1 Introduction

A phage is a type of virus that parasitizes prokaryotic microorganisms such as bacteria and archaea. Phages are the most diverse and abundant biological entities on Earth, with a total number of up to 10^{31} particles (10 times the number of bacterial particles), which are widely found in water, soil, animals, the human gut, and fermented foods (Ledormand et al., 2020; Brown et al., 2022). In the fermented food industry, phages are often of great concern due to their ability to cause fermentation failure by lysing the starter cultures (Chen et al., 2016). Currently, numerous phages have been isolated from various fermented foods through traditional culture-dependent methods, such as sauerkraut, natto, fermented cucumber, and wine (Kelsey et al., 2022). Metagenomic sequencing and data are also being used to uncover the composition of the phages in fermented foods (Park et al., 2011; Dugat-Bony et al., 2020; Du et al., 2023). However, these studies have mostly focused on the composition and diversity of phages, while there is limited research on the functional potential of phages in fermented foods (Yu et al., 2022; Du et al., 2023). To date, only one study has reported the interaction between phages and hosts (metagenome-assembled genome, MAG) based on shotgun metagenomics (Queiroz et al., 2023).

Phages, as accessory gene reservoirs, can affect the function of microbial communities through the incorporation and expression of auxiliary metabolic genes (AMGs) in ways that go beyond killing their hosts. AMGs are host-derived genes present in phage genomes that may encode proteins to redirect and enhance the metabolism of the infected host (Hayes et al., 2017). Studies in natural ecosystems have shown that AMGs possess a range of metabolic functions, including antioxidation, carbon metabolism, cell protection, cycling of nutrients, and fatty acid metabolism (Breitbart et al., 2018). Recently, there have also been a small number of reports on the functional potential of phages based on the Clusters of Orthologous Groups of proteins (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and CAZymes annotation at the community level in fermented foods (Tamang et al., 2022; You et al., 2022; Du et al., 2023). However, the detailed AMGs contained in a specific phage genome are rarely disclosed (Paillet et al., 2022), which may hamper the full understanding of their auxiliary contributions to microbiome functioning in fermented foods.

Soy sauce, a traditional soy-based fermented condiment, is widely consumed in China and other Asian countries because of its special taste and flavor (Feng et al., 2013; Liu et al., 2015). It was found that the predominant microbial groups in soy sauce fermentation systems included bacterial genera such as *Tetragenococcus*, *Weissella*, *Staphylococcus*, and *Bacillus* and fungal genera such as *Aspergillus*, *Saccharomyces*, *Candida*, and *Debaryomyces* (Han et al., 2020; Qi et al., 2021; Tan et al., 2022). Compared with the bacterial and fungal communities, there is less research on the phage community during soy sauce fermentation. There are limited reports on the isolation of phages (Uchida and Kanbe, 1993; Wakinaka et al., 2022) and the mechanisms of host-phage interactions, such as specific binding receptors (Wakinaka et al., 2022). The diversity and function of phages as well as phage-host interactions during soy sauce fermentation remain largely

unknown. To address this research gap, this work characterized the diversity and composition of phages and further analyzed the phage metabolic functions and phage-host interactions in raw soy sauce with two different types of fermentations [Cantonese-type fermentation (NJ) and Japanese-type (PJ) fermentation] based on shotgun metagenomic sequencing. These results will contribute to a better understanding of the diversity and functions of phages in soy sauce fermentation ecosystems.

2 Materials and methods

2.1 Sample collection and physicochemical analysis

Raw soy sauce samples during NJ and PJ fermentation were collected from a local soy sauce factory (Pearl River Bridge Biotechnology Co., Ltd.; Zhongshan, Guangdong, China). The samples were collected on day 120 of fermentation. The samples collected from the NJ and PJ fermentation processes were named NJ5 and PJ6, respectively. Triplicate raw soy sauce samples were collected from three different tanks and placed in 50-mL centrifuge tubes and then transported on ice to the laboratory and stored at -20°C until further analysis.

The pH, total acidity (TA), NaCl, and amino acid nitrogen (AAN) contents were analyzed using the titration method with an automatic potentiometric titrator (905-Titrando; Metrohm, Switzerland), as previously described (Tan et al., 2022).

2.2 DNA extraction, metagenomic sequencing, and assembly

The total DNA from raw soy sauce samples (0.5 g) was extracted as described previously (Tan et al., 2022). A Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, United States) and electrophoresis on a 1% agarose gel were used to determine the DNA quantity and quality. DNA extracted from triplicate samples at the same sampling time were pooled and stored at -20°C for subsequent metagenomic analysis. To obtain metagenomic data, extracted DNA was sheared into ~ 350 -bp fragments using a Covaris M220 nucleic acid shearer (Covaris, Woburn, MA, USA). The total DNA was used for sequencing library preparation with the NEBNext Ultra DNA Library Prep Kit (NEB, Ipswich, MA, USA) and then sequenced on the Illumina NovaSeq 6000 platform (San Diego, CA, USA), leading to 2×150 bp, paired-end reads, at the Novogene Bioinformatics Technology (Beijing, China). Raw sequencing data were deposited in the Sequence Read Archive (SRA) of NCBI BioProject under accession number PRJNA1026665.

2.3 Identification and classification of phage scaffolds

Adapter sequences were removed from the generated reads, which were then trimmed using Trimmomatic v.0.30 with a quality cutoff of 30, a sliding window of 6 bp, and a minimum length

cutoff of 45 bp (Bolger et al., 2014). High-quality reads were assembled using MEGAHIT v1.2.9 (Li et al., 2015). Contigs \geq 2.5 kb from metagenome assemblies were used to recover phage sequences. For each phage population, the open reading frames (ORFs) were predicted using Prodigal v2.6.3 (Hyatt et al., 2010) with the parameters set as “-p meta -g 11 -f gff -q -m -c.” Phage sequences were identified using VirSorter v1.0.6 (Guo et al., 2021) with default settings and Vibrant v1.2.1 (Kieft et al., 2020). CheckV v1.0.1 was used to assess the completeness, contamination, and quality of phage genomes represented by viral scaffolds (Nayfach et al., 2021). The candidate phages were then clustered based on 95% average nucleotide identity (ANI) over 80% of the sequence length, generating 397 species-level phage populations (viral operational taxonomic units, vOTUs). Phylogenetic trees were constructed using MEGA X software (Kumar et al., 2018) by comparing terminase large subunit (*TerL*) amino acid sequences

with viral reference sequences and NCBI-NR by BLASTP. The resulting trees were visualized using iTOL (Letunic and Bork, 2019). The alpha-diversity indices of phage communities were calculated using QIIME2.

2.4 Functional annotations and identification of AMGs

The predicted ORFs of phage genomes were annotated by searching against eggNOG v5.0.0 (Huerta-Cepas et al., 2019) using emapper v2.1.12 (Huerta-Cepas et al., 2017) and against KEGG (Kanehisa and Goto, 2000) using KofamScan v1.3.0 (Aramaki et al., 2020). The phage genomes were then subjected to AMG identification and genome annotation using the DRAM-v module

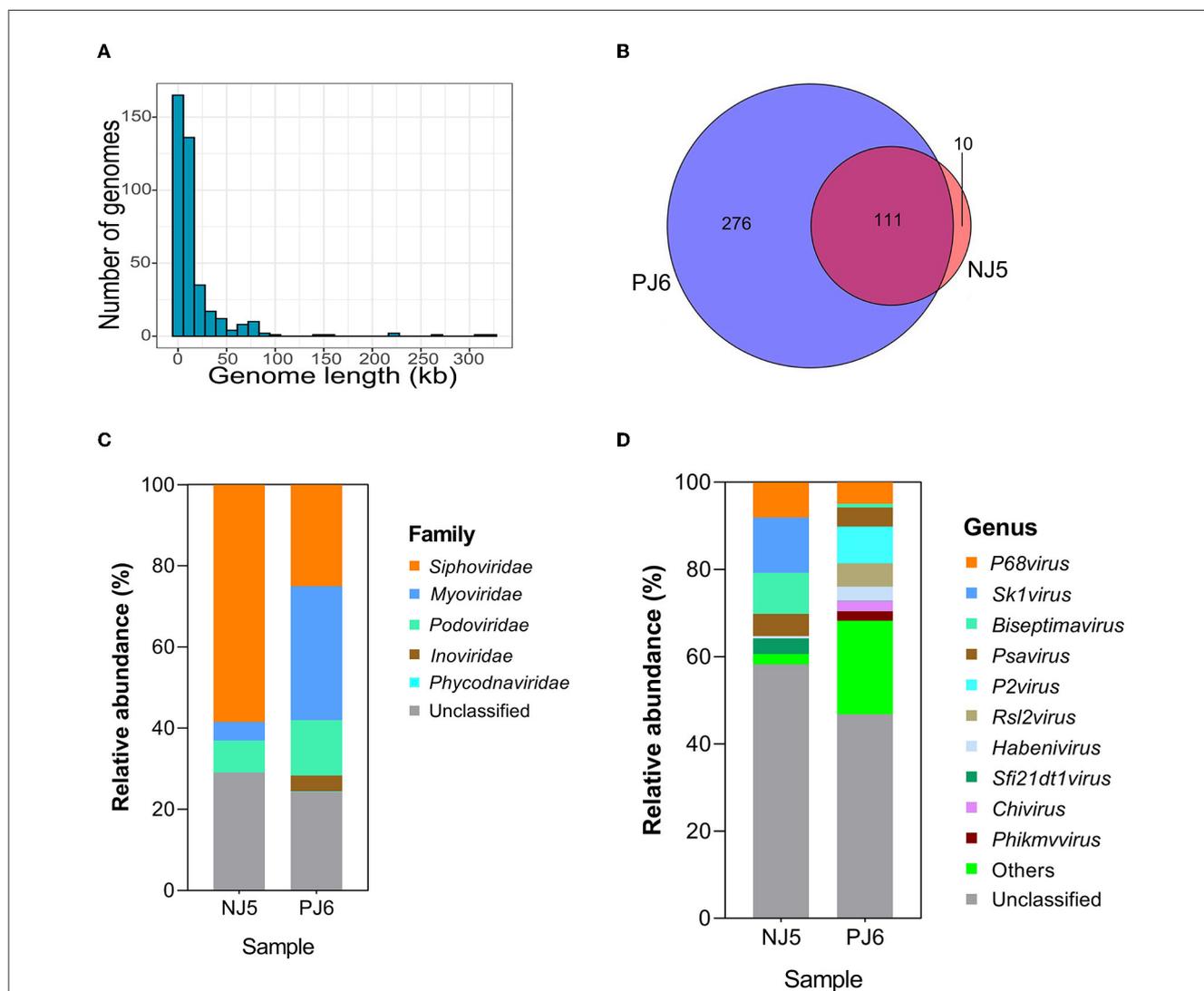


FIGURE 1 Genome size (A), Venn diagram (B), and taxonomic composition of phages at the family level (C) and genus level (D) in raw soy sauce. The Venn diagram illustrates the unique and shared viral operational taxonomic units (vOTUs) in the two samples. The 10 most abundant genera are shown. “Others” comprise the less-abundant genera. Sequences that could not be matched to any known taxonomic groups are designated as “unclassified” at the family and genus levels. NJ, Cantonese-type fermentation; PJ, Japanese-type fermentation.

from DRAM v1.4.6 (Shaffer et al., 2020). The vOTUs were searched for AMGs using the UniRef90 database and other default databases. The DRAM output contained only putative AMGs; no viral flag (F), transposon flag (T), viral-like peptidase (P), or attachment flag (A) were included, and all putative AMGs without a gene ID or gene description were also excluded. Gene arrow maps in the AMG-carrying phages were drawn using the R package gggenes (<https://cran.r-project.org/web/packages/gggenes>).

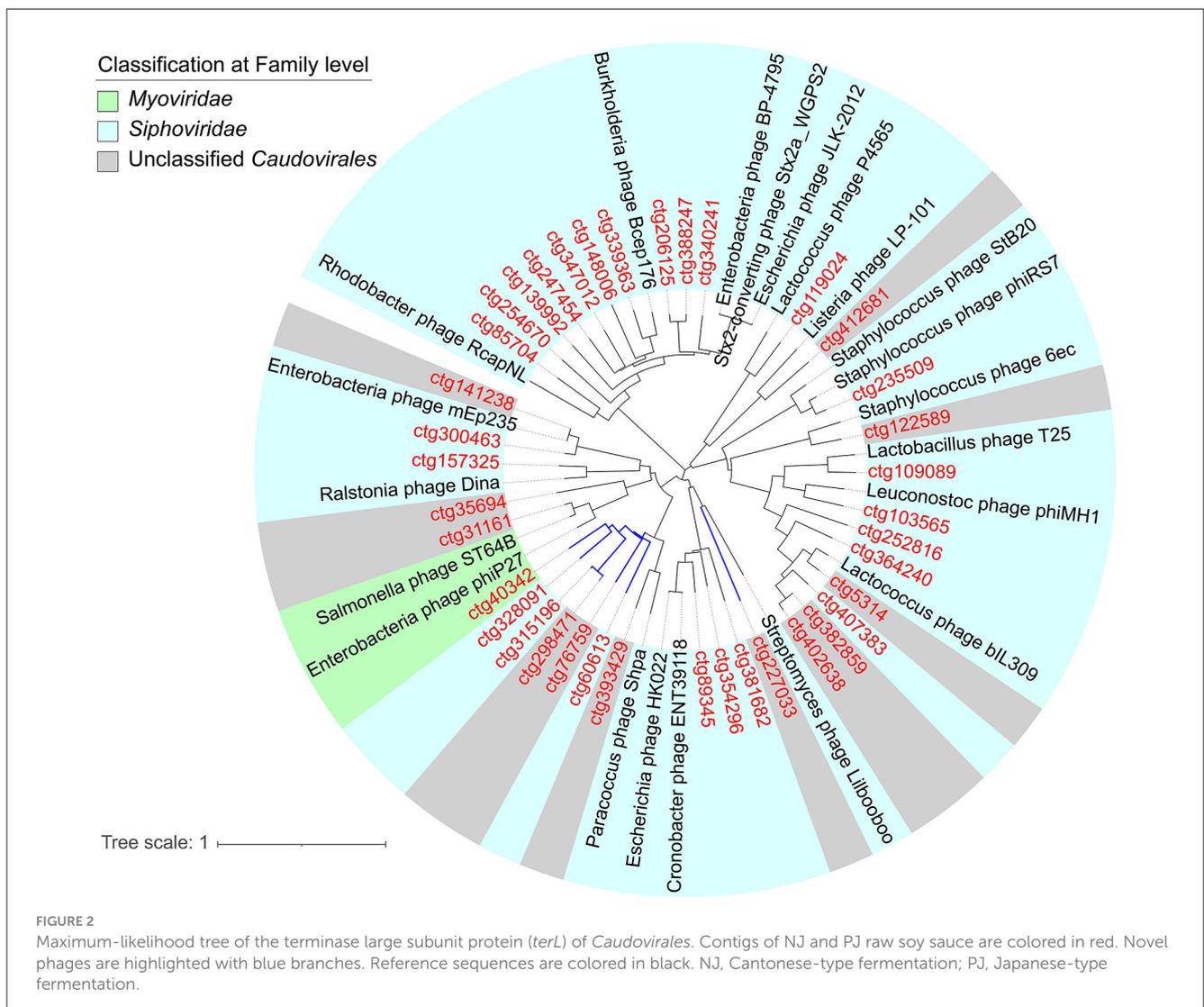
2.5 Reconstruction of host genomes

After the high-quality reads were assembled, as described previously, the assembled-contig coverage was calculated by mapping the initial reads to the contigs larger than 1 kb using the Bowtie2 software (Langmead and Salzberg, 2012). Contigs with a short length and low coverage (length < 1 kb and average coverage < 5) were discarded before the next step of binning. MetaWRAP (Uritskiy et al., 2018) was used to bin the contigs of each sample into MAGs. The qualities of MAGs (completeness

and contamination) were assessed with CheckM software (Parks et al., 2015). MAGs with low quality (50% completeness and/or 10% contamination) were removed from downstream analysis. The ORFs of all MAGs were determined using Prodigal (Hyatt et al., 2010). Then, the predicted ORFs of each MAG were checked based on sequence similarity to the NCBI-nr database using BLASTP, with a maximum allowed E-value of 1×10^{-10} . To obtain the species-level taxonomic annotation of each MAG, the BLASTP results were parsed using MEGAN (Huson et al., 2007). The MAG taxonomies were investigated using GTDB-Tk programs (Chaumeil et al., 2020). The ANI of MAGs to reference genomes was calculated using FastANI (Jain et al., 2018). Putatively novel MAGs were assigned as potentially novel species when the ANI output obtained using GTDB-Tk was <95% (Leech et al., 2020).

2.6 Phage–host linkage prediction

A sequence alignment-based method (nucleotide–nucleotide) was used to link phage genomes to their hosts *in silico* (Cheng et al.,



2022). All phage genomes were compared to the recovered bacterial genomes (MAGs) using BLASTN with the parameters of alignment length ≥ 2.5 kb, E-value $\leq 10^{-3}$, bit score ≥ 50 , and identity $\geq 70\%$ as previously suggested (Gao et al., 2022). Co-occurrence network analysis and visualization were performed using the interactive platform Gephi v0.9.2 software.

3 Results and discussion

3.1 Physicochemical characteristics of raw soy sauces

Two samples showed similar physicochemical properties. The pH in NJ5 and PJ6 was 5.07 ± 0.04 and 5.03 ± 0.07 , respectively. Also, NaCl (15.00 ± 1.45 g/100 mL in NJ5 and 14.48 ± 0.67 g/100 mL in PJ6) and AAN (0.91 ± 0.02 g/100 mL in NJ5; 0.94 ± 0.01 g/100 mL in PJ6) had similar contents. The TA contents in these two samples were slightly different, with PJ6 containing a higher TA concentration (1.83 ± 0.10 g/100 mL) than NJ5 (1.41 ± 0.02 g/100 mL).

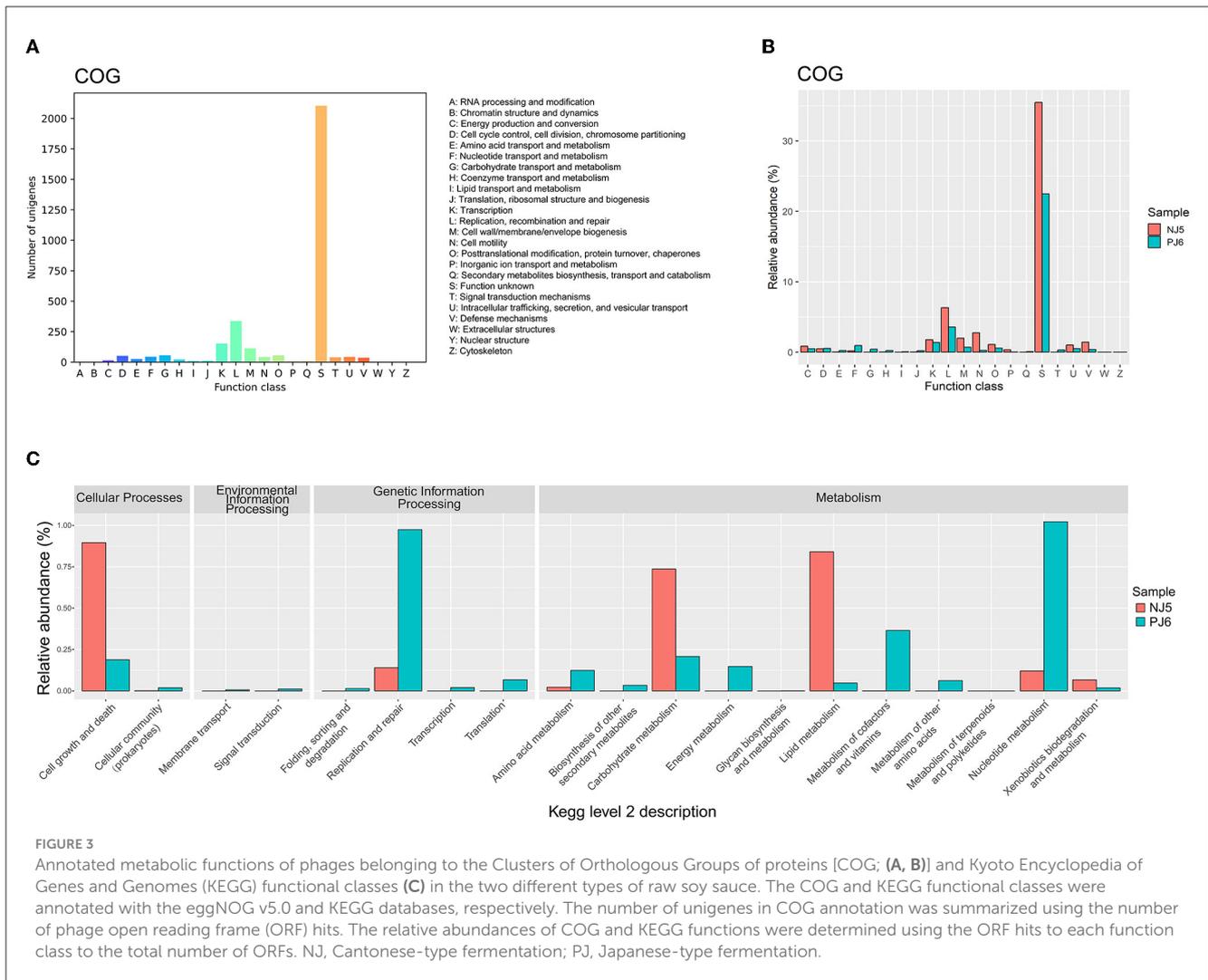
3.2 Metagenomic analysis of phage communities

The sequencing generated 70,849,140 sequences from NJ5 and PJ6 samples, with 10.2 and 9.8 Gbp of sequencing data, respectively (Supplementary Table 1). A length of 2.5 kb was selected as the minimum standard for phage genome size, and finally, 397 contigs were found with sizes ranging from 2,505 to 324,623 bp, including five complete phage genomes and six high-quality phage genomes (Supplementary Table 2). The results showed that 92.44% of contigs had lengths < 50 kb, 4.28% (20 contigs) had lengths ranging from 50 to 100 kb, and only 0.75% (three contigs) had lengths > 200 kb (Figure 1A). The size of the largest contig was 324 kb, which was a jumbo phage with a complete genome sequence. Among the 397 phages (vOTUs) detected, 111 vOTUs (28.0%) were present simultaneously in the NJ and PJ samples (Figure 1B). PJ, however, harbored a greater number of unique phages (276 out of 397, 69.5% of total phage genomes) than NJ, suggesting a high difference of phages between these two fermentation types. This work then determined the alpha diversity of the raw soy sauce phage communities and found that the Shannon index showed significant differences in phage species richness between the two groups (Supplementary Figure 1). In addition, the prophages present in the two samples had different abundances (35% in NJ and 13.7% in PJ; Supplementary Figure 2). At the family level, *Siphoviridae* (58.5%) dominated NJ, followed by *Podoviridae* (7.87%) and *Myoviridae* (4.55%), while *Myoviridae* (33.01%) and *Siphoviridae* (25.05%) were predominant families in PJ, followed by *Podoviridae* (13.63%) and *Inoviridae* (3.9%; Figure 1C). The majority of the phages in both soy sauce fermentation samples were assigned to the *Siphoviridae*, *Myoviridae*, and *Podoviridae*, belonging to tailed dsDNA *Caudovirales*. This finding was in agreement with other reports on fermented foods, such as kinema, koumiss, and kimchi (Jung et al., 2018; Kumar et al., 2019; You et al., 2022). However, the abundance of phages in the family *Inoviridae* in PJ

was remarkably higher than that in NJ, indicating the preference of these phages for particular soy sauce fermentation environments. The *Inoviridae* phages are characterized by circular ssDNA and long and thin filamentous shapes (Ishnaiwer and Al-Razem, 2013) and are usually detected at low abundances in fermented foods (Du et al., 2023). At the genus level, the two samples (NJ and PJ) exhibited different compositions, but both contained high abundances of *P68virus* and *Psavirus* (8.08 and 5.09% in NJ, respectively; 4.95 and 4.4% in PJ, respectively; Figure 1D). The relative abundances of *Sk1virus*, *Biseptimavirus*, and *Sfi21dt1* in NJ were higher than those in PJ, while *P2virus* and *Habenivirus* showed the opposite trend. In addition, *Chivirus* and *Phikmvvirus* were only detected in PJ. Subsequently, a phylogenetic tree of phages was constructed based on the *TerL* gene (Figure 2). A total of 38 TerL amino acid sequences were retrieved from contigs and used in the phylogenetic analysis. Seven novel phage contigs were discovered, one of which belonged to *Myoviridae* (ctg40342), three to *Siphoviridae* (ctg328091, ctg315196, and ctg60613), and three to unclassified *Caudovirales* (ctg298471, ctg76759, and ctg227033). The difference in the abundance of phages may be related to the relative abundance and community structure of their hosts and fermentation environmental parameters (such as temperature, oxygen availability, photoperiod, and salinity), which are different between these two soy sauce fermentation processes (Tan et al., 2022). Moreover, high abundances of unclassified phage genera (58.22% in NJ and 46.79% in PJ) were detected in the two samples, indicating that these phages are potentially novel viral species.

3.3 Potential functions of the phage community

To clarify the potential function of phages in soy sauce fermentation, this study first compared phage sequences with the eggNOG database to obtain the COG function classes. A total of 4,902 ORFs in phage sequences were finally identified. Among these ORFs, 3,172 ORFs (64.71%) were assigned to COG functions. These annotated ORFs were distributed in almost all COG functional categories (Figure 3A). The ORFs obtained were mainly clustered in “replication, recombination, and repair” (L), “transcription” (K), “cell wall/membrane/envelope biology” (M), “posttranslational modification, protein turnover, chaperones” (O), and “cell cycle control, cell division, chromosome partitioning” (D) (Supplementary Figure 3A). In addition to the conventional viral functions mentioned above, the auxiliary metabolic functions of “carbohydrate transport and metabolism” (G) and “amino acid transport and metabolism” (E) were also detected in raw soy sauce samples (NJ and PJ), indicating that phages may have a significant potential influence on the formation of soy sauce flavor. These auxiliary metabolic functions have also been found in viruses from other fermented foods (Yu et al., 2022; Du et al., 2023). During soy sauce fermentation, starch and protein substrates in soybeans are critical flavor precursors (Tan et al., 2022), which are hydrolyzed and converted into monosaccharides, peptides, and amino acids and then metabolized by microorganisms, ultimately contributing to the flavor and quality of soy sauce (Zhou et al., 2022). These carbohydrate- and amino acid metabolism-related



ARGs harbored by phages may facilitate their hosts to utilize carbohydrate and biosynthesize amino acids during soy sauce fermentation. However, most COG functions, such as “replication, recombination, and repair” (L), “cell wall/membrane/envelope biogenesis” (M), and “cell motility” (N), along with “function unknown” (S), exhibited higher abundances in NJ than in PJ (Figure 3B).

KEGG functional annotations of the phage community were performed after mapping metagenomic ORFs against KEGG databases. A total of four functional categories in KEGG level 2 were found (Supplementary Figure 3), including cellular processes, environmental information processing, genetic information processing, and metabolism. Among them, metabolic and genetic information processing was the most abundant function. However, functional differences were found at KEGG level 2 between the two samples (Figure 3C). In detail, the NJ sample exhibited a higher relative abundance of genes relating to functions in “carbohydrate metabolism,” “lipid metabolism,” and “cell growth and death,” while PJ exhibited a higher abundance of genes relating to “nucleotide metabolism,” “metabolism of cofactors and vitamins,” “amino acid metabolism,” and “replication

and repair.” The higher functional potential of amino acid metabolism identified in PJ was consistent with its COG annotation result.

After illustrating community-wide functional profiles, this work further identified the phage-encoded AMGs that were capable of modifying host metabolism during infection. Thirty AMGs were detected from 25 predicted phages, of which two AMGs (6.6%) were related to the carbohydrate-utilizing category (involved in carbohydrate incorporation), 17 (56.7%) to the MISC category (involved in pyrimidine deoxyribonucleotide biosynthesis), and 11 (36.7%) to the organic nitrogen category (involved in methionine degradation; Figure 4A). The miscellaneous function (MISC) identified in raw soy sauce was related to four enzymes: dUTP pyrophosphatase, dTMP kinase, thymidylate synthase, and dUTP pyrophosphatase, which have not been reported in fermented food systems. The phages with AMGs were selected for subsequent analysis (Figure 4B). Two phages (ctg17476 and ctg87516) that harbored *CBM32* and *CBM44*, belonging to the carbohydrate-binding module (CBM), which is important in carbohydrate metabolism, were found. The CBM family was also identified in isolated Vinitor phages

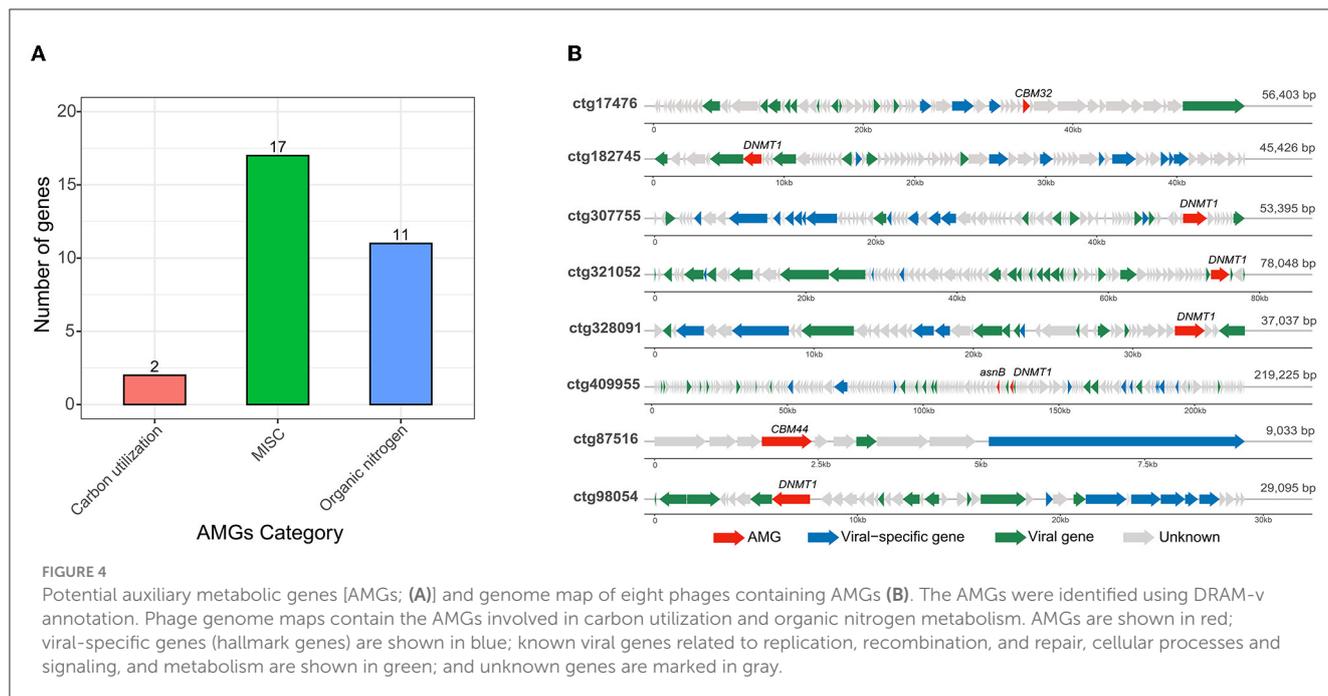


TABLE 1 Putatively novel species with FastANI identity analysis.

MAG ^a	Predicted name	Closest NCBI match ^b	Identity (%)	Closest NCBI Id	Completeness (%)	Contamination (%)
MAG9	<i>Comamonas</i> sp.	<i>Comamonas testosteroni</i> T5-67	85.81	GCF_014076415.1	84.65	2.69
MAG27	<i>Cupriavidus</i> sp.	<i>Cupriavidus</i> sp. EM10	87.48	GCF_018729255.1	84.21	4.67
MAG23	<i>Luteibacter</i> sp.	<i>Luteibacter</i> sp. UNC138MFCol5.1	85.19	GCF_900110505.1	92.74	3.68
MAG10	<i>Massilia</i> sp.	<i>Massilia</i> sp. Root133	90.45	GCF_001426525.1	81.82	1.13
MAG4	<i>Mitsuaria</i> sp.	<i>Mitsuaria</i> sp. PDC51	86.71	GCF_900113225.1	90.68	6.29
MAG32	<i>Sphingomonas</i> sp.	<i>Sphingomonas trueperi</i> DSM 7225	81.84	GCF_011927635.1	85.25	4.60

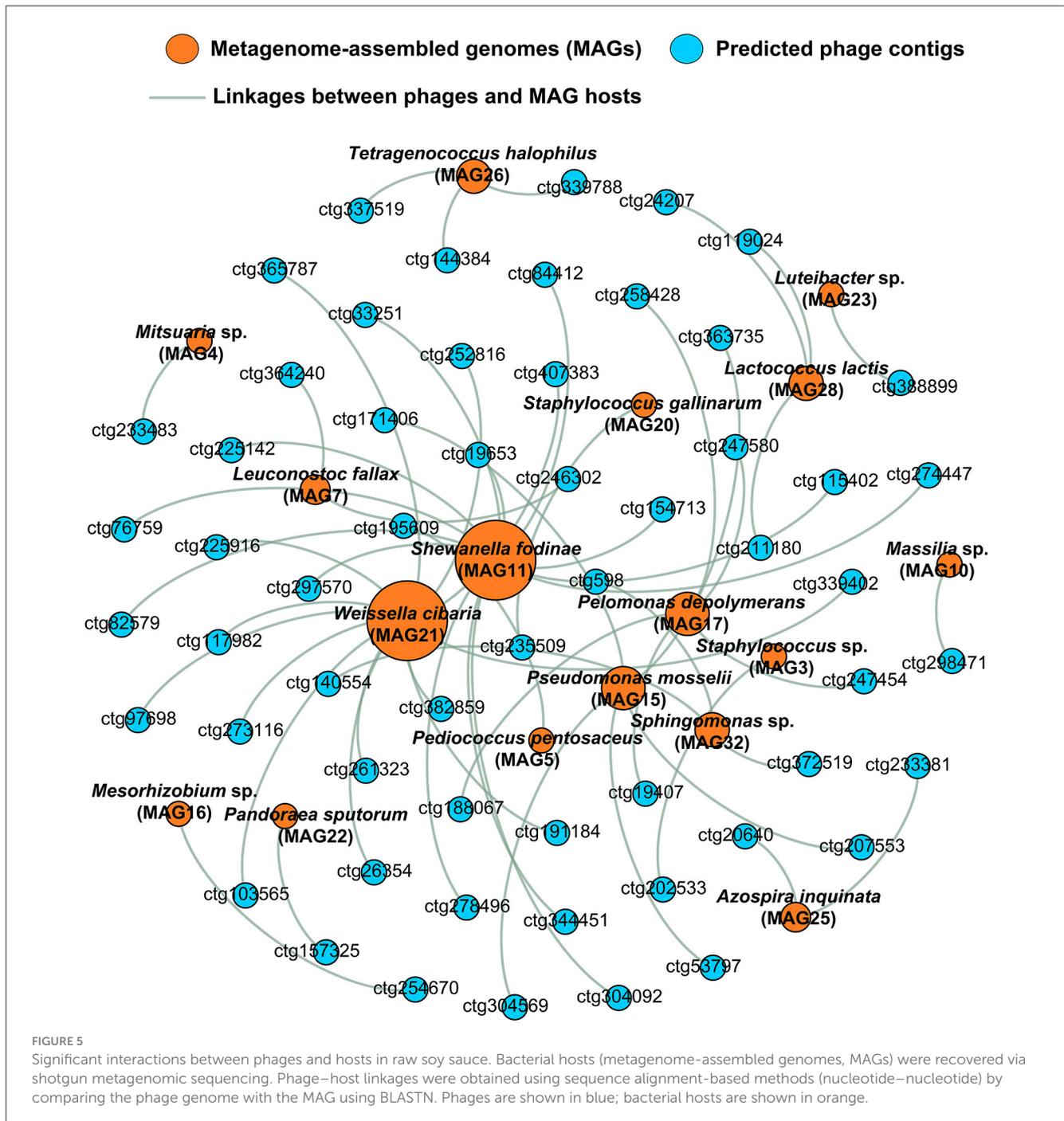
^aOnly MAGs with completeness >80% and contamination <10% have been included.

^bBased on PhyloPhlAn and MEGAN output.

from wine fermentation (vintage; Philippe et al., 2021). Six phages (ctg182745, ctg307755, ctg321052, ctg328091, ctg409955, and ctg98054) were found to possess gene *DNMT1* encoding DNA (cytosine-5)-methyltransferase 1 (EC:2.1.1.37), which was involved in organic nitrogen metabolism (cysteine and methionine metabolism). The gene *DNMT1* was also detected in phage genomes in drinking water distribution systems (Huang et al., 2023). In addition, one contig (ctg409955) was found to carry the gene *asnB* (Figure 4B), which is also associated with organic nitrogen metabolism. The gene *asnB* encodes glutamine-dependent asparagine synthetase (EC:6.3.5.4), also known as L-asparaginase II, an enzyme that converts L-asparagine to aspartic acid and ammonia (Baral et al., 2021) that is involved in alanine, aspartate, and glutamate metabolism.

3.4 Phylogenetics of MAGs reconstructed from the raw soy sauce metagenome

To reveal interactions between phages and hosts, this work reconstructed the host genomes based on metagenomic binning. This study obtained 30 strain-level MAGs with at least 50% completeness and <10% contamination (Supplementary Table 3), of which eight were characterized by high quality (>90% completeness and <5% contamination). These recruited MAGs were assigned to 22 genera, with the genus *Corynebacterium* harboring the most MAGs (four MAGs), followed by *Weissella* (three MAGs). Many species were recovered, including *Weissella cibaria* (MAG21), *Lactococcus lactis* (MAG28), and *Tetragenococcus halophilus* (MAG26), which are usually detected



in the fermentation of soy sauce (Devanthi and Gkatzionis, 2019) or other foods (Xie et al., 2018). In addition, six novel MAGs belonging to six different genera were found (Table 1). Some unexpected species, including *Azospira inquinata* (MAG25), *Shewanella fodinae* (MAG11), and *Sediminibacterium* sp. RGIG5603 (MAG19), were also reconstructed. Overall, the majority of species belonging to the dominant genus in the soy sauce fermentation system (Tan et al., 2022) were reconstructed. Some low-abundance genera such as *Stenotrophomonas*, *Enterobacter*, *Lactobacillus*, and *Citrobacter*, were not recovered by metagenomic binning, although they were identified within

the TOP15 genera by metagenomic sequencing in this study (data not shown), indicating that binning is more likely to reconstruct species with high abundance.

3.5 Interaction between phage community and host in raw soy sauce

Phage–host interactions are crucial for understanding microbial ecology and function. The genome sequence alignment-based method (nucleotide–nucleotide) is one of the important

methods used to predict the hosts of uncultivated phages (Gao et al., 2022). To our knowledge, this is the first report to predict the interaction between MAGs and phage genomes in soy sauce fermentation systems. Based on nucleotide sequence homology matches, 57 phage contigs were predicted that were associated with 17 MAGs (Figure 5; Supplementary Table 4), among which *Shewanella fodinae* (MAG11) and *Weissella cibaria* (MAG21) were the most phage-infected bacteria (each infected by 13 phages), followed by *Pseudomonas mosselii* (MAG15) and *Roseateles* sp. (MAG17; both infected by five phages). It is worth mentioning that besides *Weissella cibaria*, some species that were considered to have significant contributions to flavor compounds in soy sauce (Tan et al., 2022), including *Tetragenococcus halophilus* (MAG26), *Lactococcus lactis* (MAG28), *Pediococcus pentosaceus* (MAG5), and *Leuconostoc fallax* (MAG7), had also been infected by phages, harboring 3, 3, 1, and 2 phages, respectively. In particular, the phages of *Tetragenococcus halophilus* detected in this study were found to be phages ctg339788, ctg337519, and ctg144384 (Figure 5), which belonged to *Siphoviridae* and unclassified phages. These results were similar to those of previous reports (Higuchi et al., 1999), in which the *Siphoviridae*-related *Tetragenococcus halophilus* phages $\phi D-86$ and $\phi D10$ were also characterized. For the phage of *Tetragenococcus halophilus*, three phages (ctg339788, ctg144384, and ctg147612,) were found in this bacterium based on CRISPR-spacers match (E-value $\leq 10^{-10}$; 100% of sequence similarity over the whole spacer), in which phages ctg339788 and ctg144384 were also identified using the nucleotide-nucleotide alignment described above. However, only a very small number of hosts can be predicted based on this CRISPR-spacers match (data not shown). This might be related to the integrity of phages and MAGs (Dikareva et al., 2023). The above results suggested an interaction between a single host and multiple bacteriophages in soy sauce fermentation systems. In addition, it was found that all the phages only infected a single host, suggesting a high degree of specificity that was consistent with phages in other habitats (Achudhan et al., 2023). Here, we speculate that these phage–host relationships influence fermentation dynamics and ultimately the flavor profile of soy sauce through infecting key flavor-producing species such as *Tetragenococcus halophilus*, *Weissella*, and *Staphylococcus*.

4 Conclusion

In this study, the phage communities and phage–host interactions in raw soy sauce samples obtained from two different types of fermentation technology were investigated. The phage composition varied greatly across different fermentation processes, with over half of phage genomes detected in only the PJ sample. The comprehensive analysis of metabolic function revealed that phages encoded a large number of genes with unknown functions, and phages in the NJ sample encoded more AMGs for carbohydrate metabolism and lipid metabolism. Specifically, a large number of flavor-related bacterial species were identified as phage hosts, including *Weissella cibaria*, *Tetragenococcus halophilus*, and *Pediococcus pentosaceus*, suggesting a high level of interaction among phages and bacteria. Overall, this study provides the first preliminary characterization of the phage community diversity

and function and phage–host interactions in soy sauce. Future work is needed to better understand phage adaptability to abiotic factors, co-evolution between hosts and their phages, and how phages affect the bacterial community and flavor characteristics during the entire soy sauce fermentation process, which are critical for understanding phage–host–environment interactions in these particular high-salt niches.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://www.ncbi.nlm.nih.gov/>, PRJNA1026665.

Author contributions

GT: Project administration, Writing—original draft, Writing—review & editing, Funding acquisition, Methodology. SQ: Methodology, Writing—original draft. YW: Formal analysis, Writing—review & editing. XuL: Investigation, Writing—original draft. XiL: Software, Writing—original draft. ML: Validation, Writing—original draft. LL: Formal analysis, Writing—original draft. LZ: Resources, Writing—original draft. MH: Project administration, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Natural Science Foundation of Guangdong Province (Grant No. 2022A1515012158), the National Science Foundation of China (Grant No. 41977138), the Construction Project of Teaching Quality and Teaching Reform in Guangdong Province (Grant No. SJD202001), the General University Project of Guangdong Provincial Department of Education (Grant Nos. 2021KCXTD070 and 2021ZDZX4072), the Key Project of Social Welfare and Basic Research of Zhongshan City (Grant No. 2020B2010), and the start-up fund from the Zhongshan Institute at the University of Electronic Science and Technology in China (Grant No. 419YKQN12).

Acknowledgments

The authors are grateful to Ecogene Biotech Co., Ltd. (Shenzhen, China) for assistance with the bioinformatic analysis of the metagenomic data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

The reviewer Z-LX declared a shared affiliation with the authors LZ to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1328158/full#supplementary-material>

References

- Achudhan, A. B., Kannan, P., and Saleena, L. M. (2023). CRISPR detection in metagenome-assembled genomes (MAGs) of coal mine. *Funct. Integr. Genomics*. 23, 122. doi: 10.1007/s10142-023-01046-8
- Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2020). KofamKOALA: KEGG ortholog assignment based on profile hmm and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi: 10.1093/bioinformatics/btz859
- Baral, A., Gorkhali, R., Basnet, A., Koirala, S., and Bhattarai, H. K. (2021). Selection of the optimal l-asparaginase ii against acute lymphoblastic leukemia: an *in silico* approach. *JMIRx Med.* 2, e29844. doi: 10.2196/29844
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Breitbart, M., Bonnain, C., Malki, K., and Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* 3, 754–766. doi: 10.1038/s41564-018-0166-y
- Brown, T. L., Charity, O. J., and Adriaenssens, E. M. (2022). Ecological and functional roles of bacteriophages in contrasting environments: marine, terrestrial and human gut. *Curr. Opin. Microbiol.* 70, 102229. doi: 10.1016/j.mib.2022.102229
- Chaumeil, P., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). Gtdb-tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848
- Chen, X., Xi, Y., Zhang, H., Wang, Z., Fan, M., Liu, Y., et al. (2016). Characterization and adsorption of *Lactobacillus* virulent phage P1. *J. Dairy Sci.* 99, 6995–7001. doi: 10.3168/jds.2016-11332
- Cheng, R., Li, X., Jiang, L., Gong, L., Geslin, C., and Shao, Z. (2022). Virus diversity and interactions with hosts in deep-sea hydrothermal vents. *Microbiome* 10, 235. doi: 10.1186/s40168-022-01441-6
- Devanathi, P. V. P., and Gkatzionis, K. (2019). Soy sauce fermentation: microorganisms, aroma formation, and process modification. *Food Res. Int.* 120, 364–374. doi: 10.1016/j.foodres.2019.03.010
- Dikareva, E., Matharu, D., Lahtinen, E., Kolho, K. L., De Vos, W. M., and Salonen, A. (2023). An extended catalog of integrated prophages in the infant and adult fecal microbiome shows high prevalence of lysogeny. *Front. Microbiol.* 14, 1254535. doi: 10.3389/fmicb.2023.1254535
- Du, H., Chen, B. W., Fu, W. B., Yang, F., Lv, X. B., Tan, Y. W., et al. (2023). Composition and function of viruses in sauce-flavor baijiu fermentation. *Int. J. Food Microbiol.* 387, 110055. doi: 10.1016/j.ijfoodmicro.2022.110055
- Dugat-Bony, E., Lossouarn, J., De Paep, M., Sarthou, A., Fedala, Y., Petit, M., et al. (2020). Viral metagenomic analysis of the cheese surface: a comparative study of rapid procedures for extracting viral particles. *Food Microbiol.* 85, 103278. doi: 10.1016/j.fm.2019.103278
- Feng, Y. Z., Cui, C., Zhao, H. F., Gao, X. L., Zhao, M. M., and Sun, W. Z. (2013). Effect of *koji* fermentation on generation of volatile compounds in soy sauce production. *Int. J. Food Sci. Technol.* 48, 609–619. doi: 10.1111/ijfs.12006
- Gao, S. M., Paez-Espino, D., Li, J. T., Ai, H. X., Liang, J. L., Luo, Z. H., et al. (2022). Patterns and ecological drivers of viral communities in acid mine drainage sediments across southern china. *Nat. Commun.* 13, 2389. doi: 10.1038/s41467-022-30049-5
- Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., et al. (2021). Virsorter2: a multi-classifier, expert-guided approach to detect diverse dna and rna viruses. *Microbiome* 9, 37. doi: 10.1186/s40168-020-00990-y
- Han, D. M., Chun, B. H., Feng, T., Kim, H. M., and Jeon, C. O. (2020). Dynamics of microbial communities and metabolites in ganjiang, a traditional korean fermented soy sauce, during fermentation. *Food Microbiol.* 92, 103591. doi: 10.1016/j.fm.2020.103591
- Hayes, S., Mahony, J., Nauta, A., and van Sinderen, D. (2017). Metagenomic approaches to assess bacteriophages in various environmental niches. *Viruses* 9, 60127. doi: 10.3390/v9060127
- Higuchi, T., Uchida, K., and Abe, K. (1999). Preparation of phage-insensitive strains of *Tetragenococcus halophilus* and its application for soy sauce fermentation. *Biosci. Biotechnol. Biochem.* 63, 415–417. doi: 10.1271/bbb.63.415
- Huang, D., Yuan, M. M., Chen, J. H., Zheng, X. X., Wong, D. S., Alvarez, P. J. J., et al. (2023). The association of prokaryotic antiviral systems and symbiotic phage communities in drinking water microbiomes. *ISME Commun.* 3, 46. doi: 10.1038/s43705-023-00249-1
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5,090 organisms and 2,502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085
- Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). Megan analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107
- Hyatt, D., Chen, G., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* 11, 119. doi: 10.1186/1471-2105-11-119
- Ishnaiwer, M. M., and Al-Razem, F. (2013). Isolation and characterization of bacteriophages from laban jameed. *Food Nutr. Sci.* 04, 56–66. doi: 10.4236/fns.2013.411A008

SUPPLEMENTARY FIGURE 1

Comparison of Shannon (A) and Simpson (B) alpha diversity of phage communities in two different types of raw soy sauce. NJ, Cantonese-style fermentation; PJ, Japanese-style fermentation.

SUPPLEMENTARY FIGURE 2

Relative abundances of prophages in phage communities in two different types of raw soy sauce. NJ, Cantonese-style fermentation; PJ, Japanese-style fermentation.

SUPPLEMENTARY FIGURE 3

Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic functions of phages in raw soy sauce. Functional classes were determined at the level 2 classification of KEGG annotations using whole shotgun metagenome assembly-derived open reading frames (ORFs).

SUPPLEMENTARY TABLE 1

Overview of read statistics for the metagenomic sequence dataset.

SUPPLEMENTARY TABLE 2

Quality summary of predicted phages.

SUPPLEMENTARY TABLE 3

Characteristics of curated metagenome-assembled genomes (MAGs).

SUPPLEMENTARY TABLE 4

Phage–host linkages obtained using sequence alignment-based methods by comparing the phage genome with the metagenome-assembled genome (MAG) using BLASTN.

- Jain, C., Rodriguez, R. L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput analysis of 90 k prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9, 5114. doi: 10.1038/s41467-018-07641-9
- Jung, M., Kim, M., Yun, J., Lee, J., Kim, P. S., Lee, H., et al. (2018). Viral community predicts the geographical origin of fermented vegetable foods more precisely than bacterial community. *Food Microbiol.* 76, 319–327. doi: 10.1016/j.fm.2018.06.010
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic. Acids. Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kelsey, W., Jun-Hyeok, Y., Giovanni, E., Fabio, D. B., Arjen, N., Jennifer, M., et al. (2022). Bacteriophage-host interactions as a platform to establish the role of phages in modulating the microbial composition of fermented foods. *Microbiome Res. Rep.* 1, 3. doi: 10.20517/mrr.2021.04
- Kieft, K., Zhou, Z. C., and Anantharaman, K. (2020). Vibrant: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8, 90. doi: 10.1186/s40168-020-00867-0
- Kumar, J., Sharma, N., Kaushal, G., Samurailatpam, S., Sahoo, D., Rai, A. K., et al. (2019). Metagenomic insights into the taxonomic and functional features of kinema, a traditional fermented soybean product of sikkim himalaya. *Front. Microbiol.* 10, 1744. doi: 10.3389/fmicb.2019.01744
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). Mega x: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods.* 9, 357–359. doi: 10.1038/nmeth.1923
- Ledormand, P., Desmasures, N., and Dalmasso, M. (2020). Phage community involvement in fermented beverages: an open door to technological advances? *Crit. Rev. Food Sci. Nutr.* 61, 2911–2920. doi: 10.1080/10408398.2020.1790497
- Leech, J., Cabrera-Rubio, R., Walsh, A. M., Macori, G., Walsh, C. J., Barton, W., et al. (2020). Fermented-food metagenomics reveals substrate-associated differences in taxonomy and health-associated and antibiotic resistance determinants. *Msystems* 5, 20. doi: 10.1128/mSystems.00522-20
- Letunic, I., and Bork, P. (2019). Interactive tree of life (itol) v4: recent updates and new developments. *Nucleic. Acids. Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Li, D. H., Liu, C., Luo, R. B., Sadakane, K., and Lam, T. (2015). Megahit: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Liu, J. J., Li, D. S., Hu, Y., Wang, C., Gao, B., and Xu, N. (2015). Effect of a halophilic aromatic yeast together with *aspergillus oryzae* in koji making on the volatile compounds and quality of soy sauce moromi. *Int. J. Food Sci. Technol.* 50, 1352–1358. doi: 10.1111/ijfs.12789
- Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., and Kyrpides, N. C. (2021). Checkv assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* 39, 578–585. doi: 10.1038/s41587-020-00774-7
- Paillet, T., Lossouarn, J., Figueroa, C., Midoux, C., Rue, O., Petit, M. A., et al. (2022). Virulent phages isolated from a smear-ripened cheese are also detected in reservoirs of the cheese factory. *Viruses* 14, 81620. doi: 10.3390/v14081620
- Park, E., Kim, K., Abell, G. C. J., Kim, M., Roh, S. W., and Bae, J. (2011). Metagenomic analysis of the viral communities in fermented foods. *Appl. Environ. Microbiol.* 77, 1284–1291. doi: 10.1128/AEM.01859-10
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Philippe, C., Chaib, A., Jaomanjaka, F., Claisse, O., Lucas, P. M., Samot, J., et al. (2021). Characterization of the first virulent phage infecting *Oenococcus oeni*, the queen of the cellars. *Front. Microbiol.* 11, 596541. doi: 10.3389/fmicb.2020.596541
- Qi, Q., Huang, J., Zhou, R. Q., Yang, M. L., Zhang, L., Peng, C., et al. (2021). Characterizing microbial community and metabolites of cantonese soy sauce. *Food Biosci.* 40, 100872. doi: 10.1016/j.fbio.2020.100872
- Queiroz, L. L., Lacorte, G. A., Isidorio, W. R., Landgraf, M., de Melo, F. B. D. G., Pinto, U. M., et al. (2023). High level of interaction between phages and bacteria in an artisanal raw milk cheese microbial community. *Msystems* 8, e522–e564. doi: 10.1128/mSystems.00564-22
- Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., et al. (2020). Dram for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic. Acids. Res.* 48, 8883–8900. doi: 10.1093/nar/gkaa621
- Tamang, J. P., Das, S., Kharnaor, P., Pariyar, P., Thapa, N., Jo, S., et al. (2022). Shotgun metagenomics of *Cheonggukjang*, a fermented soybean food of Korea: community structure, predictive functionalities and amino acids profile. *Food Res. Int.* 151, 110904. doi: 10.1016/j.foodres.2021.110904
- Tan, G. L., Wang, Y., Hu, M., Li, X. Y., Li, X. L., Pan, Z. Q., et al. (2022). Comparative evaluation of the microbial diversity and metabolite profiles of Japanese-style and Cantonese-style soy sauce fermentation. *Front. Microbiol.* 13, 976206. doi: 10.3389/fmicb.2022.976206
- Uchida, K., and Kanbe, C. (1993). Occurrence of bacteriophages lytic for *Pediococcus halophilus*, a halophilic lactic-acid bacterium, in soy sauce fermentation. *J. Gen. Appl. Microbiol.* 39, 429–437. doi: 10.2323/jgam.39.429
- Uritskiy, G. V., DiRuggiero, J., and Taylor, J. (2018). Metawrap-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6, 158. doi: 10.1186/s40168-018-0541-1
- Wakinaka, T., Matsutani, M., Watanabe, J., Mogi, Y., Tokuoka, M., and Ohnishi, A. (2022). Ribitol-containing wall teichoic acid of *Tetragenococcus halophilus* is targeted by bacteriophage phiWJ7 as a binding receptor. *Microbiol. Spectr.* 10, e00336–e00322. doi: 10.1128/spectrum.00336-22
- Xie, C. Z., Zeng, H. Y., Wang, C. X., Xu, Z. M., and Qin, L. K. (2018). Volatile flavour components, microbiota and their correlations in different sufu, a chinese fermented soybean food. *J. Appl. Microbiol.* 2018, 14078. doi: 10.1111/jam.14078
- You, L. J., Yang, C. C., Jin, H., Kwok, L., Sun, Z. H., and Zhang, H. P. (2022). Metagenomic features of traditional fermented milk products. *LWT-Food Sci. Technol.* 155, 112945. doi: 10.1016/j.lwt.2021.112945
- Yu, Z., Ma, Y., Guan, Y. F., Zhu, Y. Y., Wang, K., Wang, Y. Q., et al. (2022). Metagenomics of virus diversities in solid-state brewing process of traditional chinese vinegar. *Foods* 11, 203296. doi: 10.3390/foods11203296
- Zhou, X. Y., Guo, T., Lu, Y. L., Hadiatullah, H., Li, P., Ding, K. L., et al. (2022). Effects of amino acid composition of yeast extract on the microbiota and aroma quality of fermented soy sauce. *Food Chem.* 393, 133289. doi: 10.1016/j.foodchem.2022.133289