Check for updates

OPEN ACCESS

EDITED BY Eleonora Cella, University of Central Florida, United States

REVIEWED BY Ran Wang, Capital Medical University, China Yu He, China University of Geosciences Wuhan, China Raoqiong Che, Yunnan Agricultural University, China

*CORRESPONDENCE Xiaokai Wang 🖾 wangxiaokai@01lifetech.com Yunmeng Chu 🖾 ymchu1990@gmail.com

[†]These authors have contributed equally to this work

RECEIVED 21 March 2025 ACCEPTED 22 April 2025 PUBLISHED 07 May 2025

CITATION

Liu J, Dong X, Wang X and Chu Y (2025) Characterization, diversity, and biogeochemical potential of soil viruses inhabiting in Yuncheng Salt Lake. *Front. Microbiol.* 16:1597514. doi: 10.3389/fmicb.2025.1597514

COPYRIGHT

© 2025 Liu, Dong, Wang and Chu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.

Characterization, diversity, and biogeochemical potential of soil viruses inhabiting in Yuncheng Salt Lake

Jin Liu^{1†}, Xiaoxiao Dong^{2†}, Xiaokai Wang³* and Yunmeng Chu⁴*

¹Department of Life Sciences, Yuncheng University, Yuncheng, Shanxi, China, ²National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China, ³01Life Institute, Shenzhen, Guangdong, China, ⁴Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China

Viruses play a crucial role in microbial communities and can significantly influence ecosystem processes and biogeochemical cycles by regulating the structure of these communities and the metabolic functions of their hosts. Salt lakes are recognized for harboring a diverse array of halotolerant microorganisms; however, there is limited knowledge regarding the viruses and their co-occurring hosts in these halotolerant environments. Herein, 3,362 viral operational taxonomic units (vOTUs) were recovered from Yuncheng Salt Lake soils by combining with a binning method, with less than 2% showing homology to known viruses, highlighting a significant unexplored viral diversity. Virus communities were typically specific to their habitats, exhibiting low overlap across various geographical regions and other ecological environments. It was predicted that 27.4% of viruses were associated with microbial host populations, and the majority of the predicted vOTUs are linked to dominant bacteria and archaea. Metabolic predictions revealed that 568 putative auxiliary metabolic genes (AMGs) were involved in various processes related to biosynthesis and transformation reactions associated with resource utilization within host cells. The virus-encoded AMGs indicated that these viruses influenced the metabolism of carbon, nitrogen, sulfur, and phosphorus in microorganisms, as well as their adaptation to salinity. This study highlighted the unique characteristics and potential ecological roles of soil viruses in Yuncheng Salt Lake, suggesting that these viruses may significantly influence microbial communities and the biogeochemistry of salt lake soils. These findings provide valuable insights into the diversity, function, and ecology of viruses in soils surrounding salt lakes, establishing a foundation for understanding their roles in these unique ecosystems.

KEYWORDS

salt lake, soil virus, characterization, diversity, viral auxiliary metabolic gene

1 Introduction

A salt lake is a saline water body with a salinity greater than 50 g/L. It forms through a natural process that involves the continuous influx of saline water and the influence of geological conditions, leading to the accumulation of complex substances, such as salts (Lindsay et al., 2019). Collectively, salt lakes have an estimated volume of approximately 1.04×10^{5} km³, which is comparable to the volume of freshwater lakes, estimated at around 1.25×10^{5} km³ (Huang et al., 2023). Salt lakes are distributed globally, accounting for approximately 44% of the total water volume of all lakes and 23% of the overall surface area on the planet (Huang et al., 2023). Microorganisms that can endure high salinity thrive in these environments, where elevated salt concentrations inhibit the survival of many other organisms

(Zeng et al., 2022). The complexity of these microorganisms, along with their unique metabolic processes and ability to withstand extreme conditions (such as low temperatures and high salinity), makes them valuable biological resources in extreme environments, offering significant potential for practical applications. In summary, typical salt lakes exihibit a range of gradients, including varying salinity, pH, and high altitudes (Zheng and Liu, 2009; Yang et al., 2016). These characteristics provide a unique natural laboratory for studying microbial diversity and their adaptations to environmental changes.

Yuncheng Salt Lake, located on the southern edge of Yuncheng City in Shanxi Province, China, is often called the "Dead Sea of China" (Zhang et al., 2020). It is recognized as the third-largest natural inland sodium sulfate lake in the world, situated between $110^{\circ}07'$ to $110^{\circ}50'E$ and $34^{\circ}04'$ to $34^{\circ}54'N$. The lake is rich in mirabilite, salt, and various minerals, making it an important reservoir of mineral and biological resources (Gao et al., 2021). However, rapid social and economic development has led to the overexploitation of these resources, causing substantial ecological degradation. A prolonged lack of ecological water has further contributed to the lake's shrinkage, threatening the biological populations necessary for ecological balance. In response, researchers have investigated Yuncheng Salt Lake (Zhang et al., 2019; Zeng et al., 2022), focusing on the human environment, salt-tolerant flora (Gao et al., 2021), halotolerant microorganisms (Kanik et al., 2020; Ming et al., 2020), and salt production (Zhou et al., 2019; Ming et al., 2020; Gao et al., 2021). However, significant gaps remain in the research regarding the genomic diversity of viruses in Yuncheng Salt Lake and their potential ecological roles in biogeochemical cycles. Considering the distinctive salinity conditions of Yuncheng Salt Lake, it promotes a unique diversity of microorganisms and viruses. Further exploration of the diversity and structure of viral communities, as well as their potential ecological functions in biogeochemical cycles, is essential.

Viruses, as the most abundant biological entities on Earth, are non-cellular microorganisms that rely on host cells for replication (Ji et al., 2021). They play critical roles in various ecosystems-aquatic, terrestrial, and human-associated-significantly influencing the structure and functioning of microbial communities (Paez-Espino et al., 2016). By infecting and lysing host cells, viruses alter the dynamics of microbial communities and release a large amounts of metabolic substances for other microorganisms to utilize (Suttle, 2005; Mayers et al., 2023). In addition, viruses as important carriers of horizontal gene transfer, can regulate metabolic processes in ecosystems by transferring genes that encode metabolic functions (Jansson and Wu, 2023). Researches have shown that viruses can acquire auxiliary metabolic genes (AMGs) from their hosts and transfer these genes in and out of various host cells along with the viral particles (Emerson et al., 2018; Jin et al., 2019; Mara et al., 2020). When the virus is in a lysogenic state, the expression of these hostderived genes can "hijack" or enhance host metabolic processes (Warwick-Dugdale et al., 2019), thereby contributing to ecosystem functioning by encoding enzymes essential for biogeochemical transformations. These genes can also expand the environmental niches accessible to their hosts (Sun et al., 2023). Notable examples of AMGs include genes related to photosynthesis (Millard et al., 2004; Thompson et al., 2011), as well as those involved in nucleotide metabolism and the cycling of nitrogen, phosphorus, sulfur, and various carbon metabolic processes, including methane oxidation (Sun et al., 2023).

This study aimed to investigate the characterization, geographical distribution pattems, and biogeochemical potential of viruses in salt lake soils. We analyzed 48 soil metagenomes from various sites around Yuncheng Salt Lake. The recovery of the Salt Lake viral genomes revealed unique geographical distribution characteristics, detailing the diversity of viral entities, their associations with hosts, and the presence of virus-encoded AMGs. Our findings contribute hundreds of novel viral genomes from salt lake soils to the expanding catalog of known viruses, thereby enhancing our understanding of viral diversity and functionality in extreme soil ecosystems.

2 Materials and methods

2.1 Metagenomic collection and assembly

The soil samples were collected from Yuncheng Salt Lake, which is located in Shanxi Province. Soils were collected from the southeast (SE), northeast (NE), southwest (SW), and northwest (NW) directions at sites located 15 m, 30 m, 45 m, and 60 m from the salt lake in each direction. The total DNA was extracted from soil samples using the Magnetic Soil and Stool DNA Kit (DP712, TIANGEN, China), following the manufacturer's instructions. For metagenomic sequencing, a paired-end library was constructed from the total DNA using the NEBNext Ultra DNA Library Prep Kit (NEB, USA) and sequenced on an Illumina NovaSeq 6,000 platform. Detailed information regarding the study sites and sampling methods has been documented in previous research (Zeng et al., 2022). The raw reads were trimmed and quality-filtered using Fastp v0.20.1 (Chen, 2023) to generate high-quality clean reads. Then, the clean reads were assembled into contigs using SPAdes v3.15.5 (-meta) (Bankevich et al., 2012), and short contigs smaller than 1 kb were filtered out.

2.2 Prokaryotic metagenome-assembled genome generation

Contigs from each assembly were binned with parameters --maxbin2, --metabat1, and --metabat2 in metaWRAP (Uritskiy et al., 2018), then consolidated into a final bin set with the Bin_refinement module (-c 50, -x 10). For the generated bin set, utilized dRep v2.3.2 (-comp 50, -con 10, -sa 0.95) (Olm et al., 2017) to cluster and dereplicate at a 95% average nucleotide identity (ANI) threshold, resulting in 592 species-level metagenome-assembled genomes (MAGs). Taxonomic classification was conducted with GTDB-Tk v1.6.0 (Chaumeil et al., 2020; Parks et al., 2020), and genome completeness and contamination were assessed using CheckM v1.1.3 (Parks et al., 2015) with default parameters. The dataset, which consists of 592 prokaryotic MAGs, was used to predict viral hosts.

2.3 Identification of viral sequences

The procedures and criteria for viral sequence identification were outlined as follows: (1) VIBRANT v1.2.1 with the "-virome" parameter

to enhance candidate viral sequence identification (Kieft et al., 2020); (2) VirSorter v2.2.3 (Guo et al., 2021), which classified sequences as high-confidence viruses based on a score of ≥ 0.9 or a score between 0.75 and 0.9 with at least one viral gene; and (3) CheckV v0.8.1 (Nayfach et al., 2021), which identified viral sequences based on the presence of at least one viral gene. Merged the putative viral contigs and eliminated any duplicates. Utilized vRhyme v1.1.0 (Kieft et al., 2022) with default parameters to bin the viral contigs into viral metagenome-assembled genomes (vMAGs), thereby enhancing the completeness of the viral genomes. Following established protocols, putative viruses (≥ 10 kb) were selected from the vMAGs and unbinned viral contigs. For putative viruses shorter than 10 kb, complete or circular viral sequences were retained based on assessments from CheckV, VIBRANT, and vRhyme. The viral sequences were compiled and subsequently clustered using CD-HIT v4.8.1 (parameters: \geq 95% identity and \geq 85% coverage), resulting in 3,362 viral populations, or viral operational taxonomic units (vOTUs), for downstream analyses in accordance with the previously suggested pipeline (Roux et al., 2019b). Finally, the completeness and lifestyle of the 3,362 vOTU sequences were estimated using the CheckV pipeline (Nayfach et al., 2021). Using BLASTn (Altschul et al., 1997) (E value < 1e-5, identity \geq 90%, alignment length \geq 1 kb) to compare our vOTUs with the IMG/VR v4 database (Camargo et al., 2023) to evaluate the novelty of these viruses.

2.4 Taxonomic assignments and abundance profiles of viruses

Two methods were used to assign family-level taxonomic annotations to viral sequences. The first method involved genesharing network analysis using vConTACT2 v0.9.19 (Jang et al., 2019) with Viral RefSeq database (v211). This approach enabled the assignment of the viral sequences to known taxonomic genera and families. The second method utilized a previously established majority-rules approach (Gregory et al., 2019). Specifically, we implemented the Demovir pipeline¹ with default parameters (Shkoporov et al., 2019; Liao et al., 2022) to identify sequence homologies between virus-encoded proteins to a redundant viral subset of the TrEMBL database. When more than 50% of the proteins in a viral sequence belonged to the same family, the virus was classified as a member of that family (Liao et al., 2022).

2.5 Relative abundance profiles of viruses

Clean reads of each soil sample were mapped to the 3,362 vOTUs using Bowtie2 (Langmead and Salzberg, 2012) to calculate viral relative abundance profiles. The resulting SAM files were processed with SAMtools (Li et al., 2009), and then the sorted BAM files were analyzed using CoverM v0.3.1² to generate Reads Per Kilobase per Million Mapped Reads (RPKM) profiles for all samples (Yu et al., 2024).

2.6 Phylogenetic analysis of terL gene

A phylogenetic tree illustrating the viral diversity in Salt Lake soils was constructed using the gene that encodes the terminase large subunit (*terL*), a hallmark of *Caudovirales* (tailed dsDNA phages) (Cheng et al., 2022; Chu et al., 2022). The Pfam 33.1 database (Mistry et al., 2021) were used to extract the TerL protein domain from vOTUs in this study by HMMER3 (Söding, 2005). The TerL proteins of viral RefSeq were obtained from the NCBI Virus database³ in August 2024. All TerL proteins were combined and aligned these sequences using MUSCLE v3.8.31 (Edgar, 2004), and gaps were trimmed using trimAl v.1.4 (–gappyout) (Capella-Gutiérrez et al., 2009). Used IQ-TREE v1.6.12 (Nguyen et al., 2015) with the model (–m LG + R4) (Chu et al., 2022) for phylogenetic reconstruction, and manually edited the phylogenetic tree using iTOL (Letunic and Bork, 2019).

2.7 Comparison to viruses from other data sets

We conducted a comparative analysis of viral sequences from Yuncheng Salt Lake soils against similar environments and the RefSeq database through a protein-sharing network analysis. This included saline water viruses (n = 19,709) and marine sediment viruses (n = 3,317) from the IMG/VR v4 database (Camargo et al., 2023). Used Prodigal v2.6.3 (Hyatt et al., 2010) with the "-p meta" option to predict open reading frames (ORFs) and performed all-to-all BLASTp comparisons of proteins using DIAMOND v2.0.15 (Buchfink et al., 2015). vConTACT2 v0.11.3 (Jang et al., 2019), in conjunction with the Viral RefSeq database (v211), was used to cluster viral genomes into viral clusters (VCs) based on protein-sharing networks.

2.8 Prediction of virus and host linkages

The linkages between vOTUs and putative host MAGs were established using four in silico methods, as previously reported (Emerson et al., 2018; Li et al., 2021; Chu et al., 2022). These methods were based on the following ranked criteria (Emerson et al., 2018; Luo et al., 2022): (i) CRISPR spacer matches. CRISPR spacers of putative host MAGs were extracted using CRISPRCasFinder v4.2 (Mitrofanov et al., 2021), and BLASTn was used to compared these CRISPR spacers to vOTU sequences (parameters: e-value \leq 1e-5, identity \geq 95% and mismatches \leq 2) (Emerson et al., 2018). (ii) Transfer RNA (tRNA) homology. tRNA sequences in vOTUs were retrieved using tRNAscan-SE v2.0.9 (Chan et al., 2021) and run against to MAGs via BLASTn (parameters: identity \geq 90%, coverage \geq 90%) (Fan et al., 2023). (iii) Nucleotide sequence homology. vOTUs were aligned to MAGs using BLASTn, and matches with \geq 70% identity and aligned length \geq 2 kb were retained (Roux et al., 2016). (iv) Oligonucleotide frequency (ONF) similarity. The ONF-based distances of vOTUs and MAGs were calculated using VirHostMatcher v1.0.0 (Ahlgren et al., 2017) with default parameters, and matches with d2^{*} values of ≤ 0.2 were retained (Li et al., 2021; Chu et al., 2022). If multiple hosts of a

¹ https://github.com/feargalr/Demovir

² https://github.com/wwood/CoverM

³ https://www.ncbi.nlm.nih.gov/labs/virus/

specific vOTU were predicted by multiple methods, the virus-host linkages were retained (Gios et al., 2024).

2.9 Functional profiles and identification of AMGs

Gene annotations were performed for all proteins of 3,362 vOTUs using eggNOG-mapper v2.085 (Cantalapiedra et al., 2021) with the eggNOG database v5.0 (Huerta-Cepas et al., 2019), with the parameters: e-value \leq 1e-5, bit score \geq 50. The functions of these viral genes were classified according to COG functional categories. Viral AMGs were predicted for vOTUs using two methods: (i) the DRAMv1.4.6 pipeline (Shaffer et al., 2020). VirSorter2 (--prep-for-dramv) were used for vOTU sequences to generate input files for DRAM-v. In the results of DRAM-v, AMG annotations with auxiliary scores < 4 were retained. (ii) The VIBRANT v1.2.1 pipeline was used, where vOTUs were inputted to identify AMGs and provide annotations, with default parameters. Combined the AMGs predicted based on the above pipeline, manual curations were performed to remove false positive AMGs (such as nucleotide metabolism, organic nitrogen, glycosyltransferases, and ribosomal proteins), as previously reported (Li et al., 2021; ter Horst et al., 2021; Chu et al., 2022). Putative AMGs located between or alongside two viral hallmark genes or virus-like genes were considered as high-confidence viral AMGs for further analysis (Pratama et al., 2021).

2.10 Statistical analyses

All statistical analyses were conducted using R v4.2.1. Diversity indices, including Shannon, Simpson, Chao1, and PCoA (utilizing the Bray–Curtis dissimilarity metric) were calculated with the vegan (v.2.5–7) package. Data visualizations were created in ggplot2, with Sankey plots generated via ChiPlot.⁴

3 Results and discussion

3.1 Soil viruses around salt lake

The metagenomes of Yuncheng Salt Lake soils yielded 76,847 putative viral contigs, which were analyzed using three software tools to identify viral sequences. To reconstruct longer and more complete genomes, these contigs were binned with vRhyme and resulting in 4,820 vMAGs. Then, 57,237 putative viral sequences including unbinned contigs and these vMAGs were manually filtered and clustered to approximate species-level taxonomy (Gregory et al., 2019; Ma et al., 2024). Finally, this process produced 3,362 non-redundant vOTUs, comprising 2,951 vMAGs and 411 viral contigs, each with a length \geq 10 kb or represented by circular/complete sequences. The saturation cumulative curve of Salt Lake vOTUs indicated that the viral communities present in these soils were relatively well-sampled (Figure 1A). Among the 3,362 vOTUs, 95% exceeded 10 kb, while 180 vOTUs were shorter. The lengths of the vMAGs ranged from 4.27 kb to 308 kb, with most viral sequences exceeding 10 kb, although seven circular/complete genomes fell below this threshold. Three vMAGs longer than 200 kb were classified as "huge phage" based on the presence of phage-specific genes (Al-Shayeb et al., 2020). The quality and completeness of the 3,362 vOTUs were assessed using CheckV. It classified approximately 63% into four quality tiers: complete genomes (90 vOTUs; 100% completeness), high-quality genomes (114 vOTUs; >90% completeness), medium-quality genomes (319 vOTUs; 50–90% completeness), and low-quality genomes (1,593 vOTUs; <50% completeness). The completeness of 1,246 vOTUs (37%) could not be determined (Figure 1B; Supplementary Table S1). The proportion of high-quality viral genomes in Salt Lake soils was only 6%, whereas it reached 9% in the Challenger Deep oceanic trench (Zhou et al., 2022), even without binning to improve viral genome completeness. Possible reasons for this difference include, on the one hand, the low abundance of microbial communities in salt lake soils, on the other hand, insufficient metagenomic sequencing. The completeness of the viral genomes may be improved by increasing sequencing depth or by employing HiFi metagenomic sequencing (Tao et al., 2023).

Of the 3,362 vOTUs, 79% were successfully classified (Figure 1B; Supplementary Table S1), and genome length and coverage of these identified viral sequences were shown in Figure 1C. Most soil viral sequences belonged to the order Caudovirales, accounting for 79.08% (n = 2,601) of the total. This included the families Siphoviridae (n = 1,529, 42.5%), Myoviridae (n = 465, 13.8%), and Podoviridae (n = 106, 3.2%). The remaining sequences (n = 501, 14.9%) were classified to Caudovirales but did lacked specific family classification. Caudoviricetes are characterized as double-stranded DNA (dsDNA) tailed phages, which typify most viruses that infect bacteria and archaea (Ackermann and Prangishvili, 2012). These viral families have been well-documented in deep-sea sediments (Li et al., 2021; Zhou et al., 2022; Yu et al., 2024), global soil environments (Ma et al., 2024), and saline lake ecosystems (Gu et al., 2021; Tao et al., 2023; ZeinEldin et al., 2023). Additionally, nucleocytoplasmic large DNA viruses (NCLDVs) that infect eukaryotic organisms were identified in Salt Lake soils, including *Phycodnaviridae* (n = 38, 1.13%), *Mimiviridae* (n = 10, 0.30%), and Iridoviridae (n = 3, 0.09%), along with other families such as Marseilleviridae, Ascoviridae, and Poxviridae. Lavidaviridae that are "virophages," classified as parasites of giant eukaryotic viruses (Roux et al., 2023), were also detected. The relative abundances of NCLDVs and virophages were notably low, each comprising less than 1%, consistent with previous findings in the virome of salt lakes on the Qinghai-Tibet Plateau (Gu et al., 2021). Single-stranded DNA (ssDNA) viruses (Inoviridae) and RNA viruses (Nyamiviridae) were also found in Yuncheng Salt Lake soils, these viruses have been previously identified in oceanic environments, including deep-sea sediments (Yu et al., 2024). However, their numbers and categories remain limited, and their diversity and function warrant further investigation. It should be noted that some DNA viruses may contain RNA virus marker genes, such as the RNA-Dependent RNA Polymerase (RdRp) gene, which can result in their misclassification as RNA viruses. Consequently, few RNA viruses may be detected in DNA-based metagenomic data. Yu et al. reported the identification of several RNA virus families within deep-sea metagenomic data (Yu et al., 2024).

To assess the lifestyle of viral sequences in Salt Lake soils, potential temperate viruses typically present as prophages or viral sequences

⁴ https://www.chiplot.online/



within integrase-encoding genes, were predicted from all vOTUs using VIBRANT with the KEGG, Pfam, and VOG databases (Kieft et al., 2020). Our findings revealed that 2,365 vOTUs, representing 70% of the viral communities in Salt Lake soil, could not be classified as either lytic or lysogenic (see Figure 1B; Supplementary Table S1). A significant proportion of these unclassified viral contigs may be associated with viruses exhibiting a lytic lifestyle in Yuncheng Salt Lake (Wang and Li, 2018; Wang et al., 2022). This observation aligns with previous studies of viral communities in saline lakes on the Qinghai-Tibet Plateau, which reported elevated rates of viral lysis (Gu et al., 2021). Lysogeny was predicted in only 8% (273 out of 3,362) of the Salt Lake soil viruses, contrasting with sediments of Baltic Sea (approximately 19%), where lysogeny appears more prevalent (Cai et al., 2019). Nevertheless, this finding is consistent with predictions from deep-sea sediments (~7%) (Zhou et al., 2022; Yu et al., 2024) and saline soil viruses (7%) (Gu et al., 2021) as determined by VIBRANT. It is important to interpret our conclusions cautiously, given that 70% of the viral contigs remain unclassified as either lytic or lysogenic.

3.2 Novelty of salt lake soil viruses

The majority of viruses identified in Yuncheng Salt Lake soils had not been uncharacterized previously. Specifically, consistent with prior reports (Gu et al., 2021), analysis showed that 98.69% (3,318 out of 3,362) of the vOTUs were novel, with no homologs in the IMG/VR 4 dataset, which includes various habitats (Camargo et al., 2023). Of the 44 viral sequences with homologs, 4 were classified as unclassified *Caudovirales*, and 4 had no taxonomic classification. These results highlight the distinctive characteristics of the virome observed in this study, as well as the limitations inherent in virome research conducted in salt lakes.

The *terL* gene, as a hallmark that is frequently used for phylogenetic reconstruction of *Caudovirales* phages and appears to be the best phylogenetic marker due to its widespread presence among phages and its high level of sequence conservation (Low et al., 2019; Barylski et al., 2020). Therefore, we constructed phylogenetic trees using TerL protein sequences to assess the evolutionary distance

among bacteriophages in Salt Lake soils. We identified 413 complete ORFs encoding TerL proteins in 413 viruses (413/3362 vOTUs, 12.28%) from Salt Lake soils, and then combined these with TerL proteins from viral references to construct a phylogenetic tree (Figure 2). As shown in Figure 2, the evolutionary branches of soil viruses are widely distributed across the phylogenetic tree, with the majority classified within the family *Siphoviridae* (n = 598), followed by Myoviridae (n = 94) and Podoviridae (n = 73). The novel soil viruses associated with Yuncheng Salt Lake exhibit significant diversity, as evidenced by the multiple independent small branches formed by the identified viral sequences. These clades represent previously uncharacterized diversity and may be considered novel branches of the Caudovirales, highlighting the genomic diversity of phages in our virome. The phylogenetic distances observed among viruses within a single clade reflect a comparable evolutionary diversity of Caudovirales across all soil samples. The close phylogenetic relationships among these clades and the viruses identified in soils provide insights into potential host associations, which may extend to other unclassified Caudovirales.

In addition, vConTACT2 utilizes gene-sharing networks to cluster viral genomes into VCs at approximately the genus level. This approach allows for the exploration of relationships between vOTUs in Salt Lake soil and publicly available viral genomes from various ecosystems. The analysis identified a total of 5,358 VCs from vOTUs in Salt Lake soil, saline water, marine sediment, and Viral RefSeq, with only three VCs common across all ecosystems (Figure 3; Supplementary Table S2). The limited clustering among viral genomes suggests significant habitat specificity. In Salt Lake soil, 1,475 out of 3,362 vOTUs were clustered into 672 VCs, with 56.13% of which were not found in other ecosystems, indicating a considerable proportion of endemic viruses. Additionally, only nine VCs (43 vOTUs) were shared with saline water, five VCs (5 vOTUs) with marine sediment, and eight VCs (9 vOTUs) with viral RefSeq (Figure 3; Supplementary Table S3). These results suggest that Yuncheng Salt Lake likely harbors numerous unique viruses not identified in other environments, due to its distinct geographical location and saline conditions.

3.3 Distribution and abundance of viral communities

The abundance of vOTUs varies significantly by direction and distance from Salt Lake. Samples collected from 60 m accounted for 34.2% of the total vOTUs, while those from 15 m made up 29.12%. Notably, vOTUs in the NE direction represented nearly half of total at 48.3% (Figure 4A). The distribution ratios of vOTUs across families, based on direction and distance, were shown in Supplementary Figures S2, S3. Despite proximity to Salt Lake, there were limited overlap in viral clusters, indicating distinct geographical



FIGURE 2

Phylogenetic tree of the Caudovirales viruses in soils. A maximum-likelihood tree was constructed using TerL proteins of Caudovirales. Clades and the names are colored according to viral family. The reference genomes and Salt Lake soil viruses (this study) are presented in soild pentacle and hollow pentacle, respectively. Lifestyle and quality of viruses are in presented in circles and triangles, respectively. The outermost ring show the viral host phyla (classes for Proteobacteria)



patterns in the surrounding Salt Lake soil. The number of viral clusters exclusive to the NE direction at 15 m was significantly higher than in other categories (Figures 4B,C).

To compare viral relative abundance profiles across samples, we calculated RPKM values for viral sequences classified at the family level (Figure 5A; Supplementary Table S4). Significant variations in the viral community composition of soils were observed based on directional and distance factors. Consistent with other studies on soil viromes, a substantial portion of the viral communities in Salt Lake soil comprised unclassified viruses, accounting for up to 56.81% of total viral reads (Figure 5A). In the order Caudovirales, Siphoviridae was the most dominant family, accounting for approximately 39% of the total, followed by Myoviridae at around 18% and Podoviridae at approximately 1.5%. NCLDVs made up 1.23% of the total viral abundance in Salt Lake soils, with Phycodnaviridae (0.72%) being the most abundant family, followed by Mimiviridae (0.24%). The abundance distribution of viruses within the Phycodnaviridae and Mimiviridae families showed a similar trend at the 15 m and 30 m sampling sites, but this trend reversed at the 45 m and 60 m sites. NCLDVs from these families have frequently been found in various components of ecosystems, including soil, water, ocean, and even human (Cheng et al., 2022; Schulz et al., 2022; Liang et al., 2024), but their ecological roles remain largely unexplored. Additionally, other dsDNA viruses, single-stranded DNA (ssDNA) viruses, and RNA viruses were also identified, with variations in their relative abundances across different sampling sites.

Additionally, principal coordinates analysis (PCoA) was employed to assess the degree of separation among viral communities, aiming to investigate the potential influence of geographic patterns on the distribution of viruses in Salt Lake soils. Distinct separation in viral taxonomic structure was observed in Salt Lake soil samples, with consistent distributions within samples collected from the same site (Figure 5B). Multivariate analysis methods separated samples according to various directions (Supplementary Tables S9, S10) and distances (Supplementary Tables S11, S12) (Adonis, p < 0.001). Dominant viral populations varied by distance; samples from 60 m were in the second quadrant, while others were spread across different quadrants. No clear directional distribution was observed, likely due to geographical isolation and nutrient availability variations in soils at different distances from the salt lake. The differences in viral communities observed at distinct sites in this study may be attributed to variations in in situ nutrient availability and/or differences in DNA recovery. Alternatively, these differences may also arise from variations in methods for metagenomic assembly and viral sequence extraction. In each of the four cardinal directions, the diversity scores at varying distances were corroborated by all three indices (Shannon, Simpson, and Chao1), as well as by the identified vOTUs, which were distinctly categorized based on distance (Figures 5C-E). In the northeast direction, the viral community diversity at 60 m was significantly greater than that observed in other soils, whereas in the northwest and southwest directions, the highest viral community diversity was recorded at the 30 m and 15 m sampling sites, respectively. The lowest viral diversity was contingent upon the sampling direction, with the minimum diversity recorded at 15 m in the northwest direction, 45 m in the southwest direction, and 60 m in the southeast direction. Furthermore, we aggregated all samples from various directions and distances to calculate the overall diversity index (Supplementary Figures S4, S5). The results indicate no significant differences in viral diversity across different distances, although the diversity in the northeast direction is slightly higher than in the other directions. Overall, our findings suggest that viral diversity is primarily characterized by sampling locations and is influenced by both distance and direction, with the extent of distance's impact on viral diversity varying across different directions.

3.4 Virus-host linkages in salt lake soil

Viruses can significantly shape the diversity and structure of microbial communities by transforming lifestyle, lysing host cells, or coexisting with them. Which is crucial for comprehending the ecology and function of microbial communities (Suttle, 2007). To explore potential virus-host relationships, we correlated 237 MAGs with 3,362 vOTUs using four in silico methods (Paez-Espino et al., 2016). This effort successfully linked 921 vOTUs (27.4%) with 217 MAGs, forming virus-host pairs (Figure 5A; Supplementary Table S6). This proportion aligns with findings from other environments using similar methods (Emerson et al., 2018; Li et al., 2021). The prokaryotic hosts linked to vOTUs were classified into 14 phyla, with the majority of MAGs belonging to Actinobacteriota (29.03%, 63 of 217 MAGs), followed by Proteobacteria (14.75%, 32 of 217 MAGs) and Gemmatimonadota (13.36%, 29 of 217 MAGs). Actinobacteriota (376 out of 921 vOTUs) and Gemmatimonadota (121 out of 921 vOTUs) are the most frequently predicted host phyla (Figure 6A). The vOTUs with assigned host information predominantly belong to the class Caudoviricetes (Figure 6B). Linking virus-host taxonomy revealed that viruses within the same order or family often exhibit a broad host range. Notably, approximately 92.94% of the vOTUs were predicted to target specific hosts, with only 61 vOTUs associated with hosts from different prokaryotic phyla. This included 4 vOTUs spanning 3 phyla and 57 vOTUs spanning 2 phyla. The specificity of various viruses can result in significant differences in their host ranges. Viruses with a narrow host range are restricted to specific species or strains (Holmfeldt et al., 2007), whereas those with a broad host range can infect hosts across different kingdoms and may even simultaneously infect bacteria, archaea, or fungi, such as the generalist phages (Roux et al.,



2019a). Additionally, multiple different viruses can co-infect the same host cell, as previously observed in marine bacteria (Roux et al., 2014). This observation supports existing theories and prior research indicating that most viruses exhibit a restricted host range.

3.5 Functional profiles and auxiliary metabolic genes of viruses

To investigate the potential functions of the virus in Salt Lake soil, we utilized the eggNOG database to assign functions to all 121,019 predicted genes in the viruses (Supplementary Table S7). As a result, 64,134 genes from the total viral gene pool were annotated to known orthologous groups, with 23.3% (n = 14,928) classified as functionally unknown (Figure 7A). Among the identified functional genes, a significant proportion is associated with viral replication and transcription, exhibiting high relative abundance, particularly within the categories of "replication, recombination and repair (L)," "transcription (K)" and "cell wall/ membrane/envelope biogenesis (M)." Notebly, 1.9% (n = 1,196) of the all annotated functional genes in viruses were linked to "carbohydrate transport and metabolism (G)" (Figure 7A), a finding also observed in mangrove sediment (Jin et al., 2019) and plateau salt lakes (Gu et al., 2021).

Viruses that carry host-derived AMGs can express these genes within the host cells after infection, thereby manipulating the metabolic processes of the host (Breitbart et al., 2007). Due to environmental selective pressures, those AMGs that help maintain and/or enhance the adaptability of the viruses or their hosts are retained in the genomic sequence (Breitbart et al., 2018). To elucidate the potential roles of viruses in influencing critical biogeochemical cycles, a total of 568 viral AMGs within 316 vOTUs, representing 9.4% of all vOTUs, were identified across 11 pathways as defined by KEGG that are linked to targeted processes, including carbon, nitrogen, sulfur, and phosphorus cycles. This identification was achieved through the application of DRAM-v and VIBRANT (Supplementary Table S8). Among these, the number of AMGs carried by viruses varied from 1 to 14, with the vast majority of viral sequences containing fewer than 4 AMGs (Supplementary Table S8). The viruses predominantly encoded AMGs related to cofactor and vitamin metabolism (a.g.



pantothenate and CoA biosynthesis, and folate biosynthesis intermediates), followed by amino acid metabolism (e.g., methyltransferases, aminotransferases, asparagine synthase), and carbohydrate metabolism (primarily involving the metabolism of various nucleotide, sugars and oligosaccharides) (Figure 7B; Supplementary Table S8).

Through the analysis of the composition and abundance of viral AMGs in Salt Lake soil samples, the results indicated that viruses in these soils harbored a significant number of AMGs with diverse functional roles (Figures 7C,D). Notably, AMGs associated with cysteine and methionine biosynthesis, particularly DNA [cytosine-5]-methyltransferase 1 (*dmc*) were found to be overrepresented in the viral genomes of the soil samples when

compared to other metabolic pathways. These metabolic functions have been documented as being enriched in marine viruses (Enav et al., 2014), suggesting that the selective pressures present in various habitats can result in a consistent acquisition of AMGs. Under specific environmental conditions, AMGs that entered host cells through viruses might compensate for the host's metabolic deficiencies or introduce new metabolic pathways. This alteration of the host's metabolic processes enabled the host to adapt to the current environment, as observed in marine ecosystems (He et al., 2017). After the virus carrying AMGs entered the host cell, certain AMGs could redirect the host's metabolism to enhance resource supply or recycle intracellular nucleotides for viral replication (Lavigne et al., 2013).



Additionally, some AMGs may provide the host with defense mechanisms for protection, while others, such as DNA methyltransferases, can help the virus evade host restriction endonucleases (Murphy et al., 2013).

Yuncheng Salt Lake is characterized by the presence of sodium sulfate (Na_2SO_4) (Zhou et al., 2019). Within this ecosystem, both sulfur oxidation and sulfate reduction played crucial roles in the sulfur cycling process (Liu et al., 2022). Among the 568 viral AMGs identified, 21 were annotated as being involved in sulfur metabolism. Notably, eight of these AMGs encode phosphoadenosine phosphosulfate reductases

(*cysH*), which are integral to assimilatory sulfate reduction. Recent studies have also reported the presence of AMGs coding for *cysH* in viruses obtained from deep-sea sediments (Zhou et al., 2022; Yu et al., 2024), salt lakes (Gu et al., 2021), and seawater (Li et al., 2021; Zhou et al., 2022). The remaining 13 viral AMGs were associated with the sulfur relay system, which included sulfur transferases such as sulfate adenylyltransferase (*cysN*) and sulfur-carrier protein adenylyltransferase (*moeZR*), as well as sulfur-carrier proteins like *moaD* and *mec* (CysO sulfur-carrier protein-S-L-cysteine hydrolase). Furthermore, it has been proposed that the sulfur relay system contributed to microbial



(2) (Secondary metabolites biosynthesis, transport and catabolism), 'D' (Cell cycle control, cell division, chromosome partitioning), 'U' (Intracellular trafficking, secretion, and vesicular transport), 'N' (Cell motility), 'A' (RNA processing and modification), 'B' (Chromatin structure and dynamics), 'Z' (Cytoskeleton). (B) The number of viral AMGs involved in different metabolic categories. (C) Relative abundance of the viral AMGs across Salt Lake soil samples. KEGG metabolic categories are colored according to the legend.

tolerance against acid stress (Cao et al., 2019), heavy metals (Gang et al., 2019), and organic solvents (Liu et al., 2012). Consequently, these AMGs may enhance the adaptability of their hosts to various stress conditions, thereby providing a competitive advantage in salt soil environments characterized by significant fluctuations in temperature and chemistry. Additionally, viral AMGs that facilitated the metabolism of carbon (C), nitrogen (N), and phosphorus (P) were also identified in Salt Lake soil viruses. In terms of carbon metabolism, one viral AMG encoding pyruvate phosphate dikinase (ppdk) was discovered, which was involved in the reductive tricarboxylic acid (rTCA) and dicarboxylate/4-hydroxybutyrate (DC/4-HB) cycles (Hügler and Sievert, 2011), represented the primary pathways for dark carbon fixation by deep-sea microbes (Hügler and Sievert, 2011). Regarding nitrogen metabolism, virus-encoded nirA (ferredoxinnitrite reductase), associated with dissimilatory nitrate reduction, along with numerous genes related to organic degradation and synthesis, were also detected. We also found several

virus-encoded phosphorus metabolism AMGs, including phoD, gph and pgp genes related to organic phosphoester degradation. In particular, organophosphates are recognized as emerging contaminants that pose a significant risk to human health (Silva and Baltrusaitis, 2021). The presence of viruses carrying phosphorus metabolism AMGs underscored the essential role of viruses in phosphorus metabolism. Moreover, the most abundant AMG identified in the metagenomes of Salt Lake soil was DNA cytosine methyltransferase (DNMT1, dcm). However, virusencoded dcm genes have been found in metagenomes from various environments, suggesting that they performed central functions in the viral life cycle (Waite et al., 2018). Consequently, the presence of viral AMGs that directly mediated the metabolic processes of host cells might also exert an indirect influence on biogeochemical cycles (Zimmerman et al., 2020). This suggests that viruses played a crucial role in regulating the metabolism of host cells and, in turn, affected biogeochemical cycles within this particular region.

4 Conclusion

Salt lakes harbored a diverse array of unique microorganisms and their associated viruses; however, the role of viruses in these ecosystems-particularly in their interactions with microbial hosts regarding aspects such as microbial mortality, ecological functions, and co-evolution-has largely remained underexplored. This study revealed the novelty, richness, and diversity of prokaryotic viruses harbored in soils collected from Yuncheng Salt Lake, as determined through metagenomic analysis. The viral communities exhibited distinct spatial separation in the soils surrounding Yuncheng Salt Lake. This phenomenon appeared to be influenced by variations in salt ion concentration and nutrient availability, which, in turn, shaped the structure of the prokaryotic host communities across different sampling sites. From the same metagenomic data, the identified viral sequences were in silico associated with the reconstructed prokaryotic MAGs, and a significant proportion of vOTUs could be linked to the host MAGs. It is worth noting that many of the putative soil viral microbial hosts belong to taxonomic groups that do not have cultured representatives. These findings enhanced our understanding of the diversity of halotolerant archaeal viruses, particularly those that infected important archaeal lineages present in halotolerant microorganism, such as Halobacteriota, Nanoarchaeota, and Thermoproteota. The potential presence of lytic viral communities in the salt lake soils might contribute to the high microbial density observed in this otherwise nutrient-deficient environment through the viral shunt mechanism, which influenced microbial mortality and biomass turnover. In addition, 316 soil viruses carried 568 AMGs that performed various functions related to viral replication, as well as participated in the elemental cycling of carbon, sulfur, phosphorus, and other elements. These genes may compensate for the host's metabolic pathways during viral infections or introduce new metabolic pathways to enhance the metabolic processes of prokaryotic hosts, potentially impacting the biogeochemical processes within microbial communities. Salt lakes served as natural reservoirs of halotolerant prokaryotic diversity and were vital oases for the activity of viruses associated with these microorganisms. Considering that the data were obtained from a public database, information on the environmental factors associated with the samples is lacking. This study aimed to explore the diversity, geographical distribution, and potential ecological functions of viruses in soils surrounding Yuncheng Salt Lake. Future research on salt lake should involve re-sampling and measuring key environmental parameters such as temperature, pH, salinity, and nutrient levels. These measurements should be integrated with metatranscriptomic analysis and key gene expression studies to better understand the environmental factors influencing the viral community and the functions of virus-encoded AMGs in this environment.

This study did not investigate the co-evolution of viruses with microbial communities inhabiting salt lakes and their surrounding environments, particularly concerning their potential roles in horizontal gene transfer, as well as the seasonal variations in virushost interactions. These aspects necessitate further research and exploration. Future research should focus on the high-throughput cultivation of salt lake viruses to investigate the interactions between these viruses and their hosts. The findings could offer valuable insights into the metabolic potential of microorganisms and viruses in salt lakes regarding nutrient cycling, the efficiency of viral dispersal, and the dynamics of virus-host interactions. Given that only a small portion of the identified vOTUs can be classified, and many predicted to be infected belong to poorly characterized taxa, there remains a significant gap in our understanding of the factors influencing the biogeochemistry of salt lake soils—including anthropogenic influences—as well as the factors affecting the metabolism and functional adaptations of viruses and microorganisms in these environments. These topics will be important areas for future investigation.

Data availability statement

The raw reads of metagenomes were deposited in GenBank under BioProject number PRJNA861910. The original contributions of this study are included in the article/Supplementary material. Other data will be made available on request.

Author contributions

JL: Formal analysis, Writing – original draft, Methodology, Conceptualization. XD: Writing – original draft, Visualization, Software, Formal analysis. XW: Supervision, Writing – review & editing, Project administration, Validation. YC: Writing – review & editing, Project administration, Supervision, Methodology.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the horizontal cooperation project with Shenzhen 01Life Institute (#202412A001, #202112E401).

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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.2025.1597514/ full#supplementary-material

References

Ackermann, H. W., and Prangishvili, D. J. (2012). Prokaryote viruses studied by electron microscopy. Arch. Virol. 157, 1843–1849. doi: 10.1007/s00705-012-1383-y

Ahlgren, N. A., Ren, J., Lu, Y. Y., Fuhrman, J. A., and Sun, F. Z. (2017). Alignment-free \$d_2^*\$ oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res.* 45, 39–53. doi: 10.1093/nar/gkw1002

Al-Shayeb, B., Sachdeva, R., Chen, L. X., Ward, F., Munk, P., Devoto, A., et al. (2020). Clades of huge phages from across Earth's ecosystems. *Nature* 578, 425–431. doi: 10.1038/s41586-020-2007-4

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Barylski, J., Enault, F., Dutilh, B. E., Schuller, M. B., Edwards, R. A., Gillis, A., et al. (2020). Analysis of spounaviruses as a case study for the overdue reclassification of tailed phages. *Syst. Biol.* 69, 110–123. doi: 10.1093/sysbio/syz036

Breitbart, M., Bonnain, C., Malki, K., and Sawaya, N. (2018). Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* 3, 754–766. doi: 10.1038/s41564-018-0166-y

Breitbart, M., Thompson, L. R., Suttle, C. A., and Sullivan, M. B. (2007). Exploring the vast diversity of marine viruses. *Oceanography Society* 20, 135–139. doi: 10.5670/oceanog. 2007.58

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59-60. doi: 10.1038/nmeth.3176

Cai, L., Jørgensen, B. B., Suttle, C. A., He, M., Cragg, B. A., Jiao, N., et al. (2019). Active and diverse viruses persist in the deep sub-seafloor sediments over thousands of years. *ISME J.* 13, 1857–1864. doi: 10.1038/s41396-019-0397-9

Camargo, A. P., Nayfach, S., Chen, I. M. A., Palaniappan, K., Ratner, A., Chu, K., et al. (2023). IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* 51, D733–D743. doi: 10.1093/nar/gkac1037

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, Orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293

Cao, L., Wang, J., Sun, L., Kong, Z., Wu, Q., and Wang, Z. (2019). Transcriptional analysis reveals the relativity of acid tolerance and antimicrobial peptide resistance of *Salmonella. Microb. Pathog.* 136:103701. doi: 10.1016/j.micpath.2019.103701

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 49, 9077–9096. doi: 10.1093/nar/gkab688

Chaumeil, P. A., 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, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta 2:e107. doi: 10.1002/imt2.107

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

SUPPLEMENTARY FIGURE S1

Histogram showing the distribution of viral genome size.

SUPPLEMENTARY FIGURE S2 The pie chart shows the

The pie chart shows the percentage of different virus families in different directions. NE, northeast; NW, northwest; SE, southeast; SW, southwest.

SUPPLEMENTARY FIGURE S3

The pie chart shows the percentage of different virus families in different distances.

SUPPLEMENTARY FIGURE S4

The Shannon, Simpson, and Chao1 diversity indices of viruses in Salt Lake soils across various directions. NE, northeast; NW, northwest; SE, southeast; SW, southwest.

SUPPLEMENTARY FIGURE S5

The Shannon, Simpson, and Chao1 diversity indices of viruses in Salt Lake soils across various distances.

Chu, Y. M., Zhao, Z. L., Cai, L. X., and Zhang, G. Y. (2022). Viral diversity and biogeochemical potential revealed in different prawn-culture sediments by virusenriched metagenome analysis. *Environ. Res.* 210:112901. doi: 10.1016/j.envres. 2022.112901

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 1–19. doi: 10.1186/1471-2105-5-113

Emerson, J. B., Roux, S., Brum, J. R., Bolduc, B., Woodcroft, B. J., Jang, H. B., et al. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* 3, 870–880. doi: 10.1038/s41564-018-0190-y

Enav, H., Mandel-Gutfreund, Y., and Béjà, O. (2014). Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. *Microbiome* 2, 1–12. doi: 10.1186/2049-2618-2-9

Fan, X., Ji, M., Mu, D., Zeng, X., Tian, Z., Sun, K., et al. (2023). Global diversity and biogeography of DNA viral communities in activated sludge systems. *Microbiome* 11:234. doi: 10.1186/s40168-023-01672-1

Gang, H., Xiao, C., Xiao, Y., Yan, W., Bai, R., Ding, R., et al. (2019). Proteomic analysis of the reduction and resistance mechanisms of *Shewanella oneidensis* MR-1 under long-term hexavalent chromium stress. *Environ. Int.* 127, 94–102. doi: 10.1016/j.envint.2019.03.016

Gao, F., Nan, F., Feng, J., Lü, J., Liu, Q., Liu, X., et al. (2021). Transcriptome profile of *Dunaliella salina* in Yuncheng Salt Lake reveals salt-stress-related genes under different salinity stresses. *J. Oceanol. Limnol.* 39, 2336–2362. doi: 10.1007/s00343-021-0164-4

Gios, E., Mosley, O. E., Hoggard, M., and Handley, K. M. (2024). High niche specificity and host genetic diversity of groundwater viruses. *ISME J.* 18:wrae035. doi: 10.1093/ismejo/wrae035

Gregory, A. C., Zayed, A. A., Conceiçao-Neto, N., Temperton, B., Bolduc, B., Alberti, A., et al. (2019). Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177, 1109–1123.e14. doi: 10.1016/j.cell.2019.03.040

Gu, C., Liang, Y., Li, J., Shao, H., Jiang, Y., Zhou, X., et al. (2021). Saline lakes on the Qinghai-Tibet plateau harbor unique viral assemblages mediating microbial environmental adaption. *iScience* 24:103439. doi: 10.1016/j.isci.2021.103439

Guo, J. R., 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

He, T., Li, H., and Zhang, X. (2017). Deep-sea hydrothermal vent viruses compensate for microbial metabolism in virus-host interactions. *MBio* 8, e00893–e00817. doi: 10.1128/mBio.00893-17

Holmfeldt, K., Middelboe, M., Nybroe, O., and Riemann, L. J. (2007). Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobacterium* hosts. *Appl. Environ. Microbiol.* 73, 6730–6739. doi: 10.1128/AEM.01399-07

Huang, J., Yang, J., Han, M., Wang, B., Sun, X., and Jiang, H. (2023). Microbial carbon fixation and its influencing factors in saline lake water. *Sci. Total Environ.* 877:162922. doi: 10.1016/j.scitotenv.2023.162922

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Hügler, M., and Sievert, S. M. (2011). Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Annu. Rev. Mar. Sci.* 3, 261–289. doi: 10.1146/annurev-marine-120709-142712

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Jang, H. B., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* 37, 632–639. doi: 10.1038/ s41587-019-0100-8

Jansson, J. K., and Wu, R. (2023). Soil viral diversity, ecology and climate change. *Nat. Rev. Microbiol.* 21, 296–311. doi: 10.1038/s41579-022-00811-z

Ji, M., Liu, Z., Sun, K., Li, Z., Fan, X., and Li, Q. (2021). Bacteriophages in water pollution control: advantages and limitations. *Front. Environ. Sci. Eng.* 15, 84–15. doi: 10.1007/s11783-020-1378-y

Jin, M., Guo, X., Zhang, R., Qu, W., Gao, B., and Zeng, R. (2019). Diversities and potential biogeochemical impacts of mangrove soil viruses. *Microbiome* 7:58. doi: 10.1186/s40168-019-0675-9

Kanik, M., Munro-Ehrlich, M., Fernandes-Martins, M. C., Payne, D., Gianoulias, K., Keller, L., et al. (2020). Unexpected abundance and diversity of phototrophs in mats from morphologically variable microbialites in great salt Lake, Utah. *Appl. Environ. Microbiol.* 86, e00165–e00120. doi: 10.1128/AEM.00165-20

Kieft, K., Adams, A., Salamzade, R., Kalan, L., and Anantharaman, K. (2022). vRhyme enables binning of viral genomes from metagenomes. *Nucleic Acids Res.* 50:e83. doi: 10.1093/nar/gkac341

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

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lavigne, R., Lecoutere, E., Wagemans, J., Cenens, W., Aertsen, A., Schoofs, L., et al. (2013). A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus LUZ19. *MBio* 4, e00061–e00013. doi: 10.1128/mBio.00061-13

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, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, Z. X., Pan, D., Wei, G. S., Pi, W. L., Zhang, C. W., Wang, J. H., et al. (2021). Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. *ISME J*. 15, 2366–2378. doi: 10.1038/s41396-021-00932-y

Liang, J. L., Feng, S. W., Jia, P., Lu, J. L., Yi, X., Gao, S. M., et al. (2024). Unraveling the habitat preferences, ecological drivers, potential hosts, and auxiliary metabolism of soil giant viruses across China. *Microbiome* 12:136. doi: 10.1186/s40168-024-01851-8

Liao, H., Li, H., Duan, C. S., Zhou, X. Y., Luo, Q. P., An, X. L., et al. (2022). Response of soil viral communities to land use changes. *Nat. Commun.* 13:6027. doi: 10.1038/s41467-022-33771-2

Lindsay, M. R., Johnston, R. E., Baxter, B. K., and Boyd, E. S. (2019). Effects of salinity on microbialite-associated production in Great Salt Lake, Utah. *Ecology* 100:e02611. doi: 10.1002/ecy.2611

Liu, J., Chen, L., Wang, J., Qiao, J., and Zhang, W. (2012). Proteomic analysis reveals resistance mechanism against biofuel hexane in *Synechocystis* sp. PCC 6803. *Biotechnol. Biofuels* 5, 1–17. doi: 10.1186/1754-6834-5-68

Liu, Q., Yang, J., Wang, B., Liu, W., Hua, Z., and Jiang, H. (2022). Influence of salinity on the diversity and composition of carbohydrate metabolism, nitrogen and sulfur cycling genes in lake surface sediments. *Front. Microbiol.* 13:1019010. doi: 10.3389/fmicb.2022.1019010

Low, S. J., Džunková, M., Chaumeil, P. A., Parks, D. H., and Hugenholtz, P. J. (2019). Evaluation of a concatenated protein phylogeny for classification of tailed doublestranded DNA viruses belonging to the order *Caudovirales. Nat. Microbiol.* 4, 1306–1315. doi: 10.1038/s41564-019-0448-z

Luo, X. Q., Wang, P. D., Li, J. L., Ahmad, M., Duan, L., Yin, L. Z., et al. (2022). Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. *Microbiome* 10:190. doi: 10.1186/s40168-022-01384-y

Ma, B., Wang, Y. L., Zhao, K. K., Stirling, E., Lv, X. F., Yu, Y. J., et al. (2024). Biogeographic patterns and drivers of soil viromes. *Nat. Ecol. Evol.* 8, 717–728. doi: 10.1038/s41559-024-02347-2

Mara, P., Vik, D., Pachiadaki, M. G., Suter, E. A., Poulos, B., Taylor, G. T., et al. (2020). Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline. *ISME J.* 14, 3079–3092. doi: 10.1038/s41396-020-00739-3

Mayers, K. M. J., Kuhlisch, C., Basso, J. T. R., Saltvedt, M. R., Buchan, A., and Sandaa, R. A. (2023). Grazing on marine viruses and its biogeochemical implications. *MBio* 14:e0192121. doi: 10.1128/mbio.01921-21

Millard, A., Clokie, M. R., Shub, D. A., and Mann, N. H. (2004). Genetic organization of the psbAD region in phages infecting marine *Synechococcus* strains. *Proc. Natl. Acad. Sci. USA* 101, 11007–11012. doi: 10.1073/pnas.0401478101

Ming, H., Ji, W. L., Li, M., Zhao, Z. L., Cheng, L. J., Niu, M. M., et al. (2020). *Halomonas lactosivorans* sp. nov., isolated from salt-lake sediment. *Int. J. Syst. Evol. Microbiol.* 70, 3504–3512. doi: 10.1099/ijsem.0.004209

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Mitrofanov, A., Alkhnbashi, O. S., Shmakov, S. A., Makarova, K. S., Koonin, E. V., and Backofen, R. (2021). CRISPRidentify: identification of CRISPR arrays using machine learning approach. *Nucleic Acids Res.* 49:e20. doi: 10.1093/nar/gkaa1158

Murphy, J., Mahony, J., Ainsworth, S., Nauta, A., and van Sinderen, D. J. (2013). Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. *Appl. Environ. Microbiol.* 79, 7547–7555. doi: 10.1128/AEM.02229-13

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

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Olm, M. R., Brown, C. T., Brooks, B., and Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868. doi: 10.1038/ismej.2017.126

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

Parks, D. H., Chuvochina, M., Chaumeil, P. A., Rinke, C., Mussig, A. J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* 38, 1079–1098. doi: 10.1038/s41587-020-0501-8

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

Pratama, A. A., Bolduc, B., Zayed, A. A., Zhong, Z. P., Guo, J. R., Vik, D. R., et al. (2021). Expanding standards in viromics: in silico evaluation of dsDNA viral genome identification, classification, and auxiliary metabolic gene curation. *Peerj* 9:e11447. doi: 10.7717/peerj.11447

Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., et al. (2019b). Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* 37, 29–37. doi: 10.1038/nbt.4306

Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., et al. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537, 689–693. doi: 10.1038/nature19366

Roux, S., Fischer, M. G., Hackl, T., Katz, L. A., Schulz, F., and Yutin, N. (2023). Updated *virophage* taxonomy and distinction from polinton-like viruses. *Biomol. Ther.* 13:204. doi: 10.3390/biom13020204

Roux, S., Hawley, A. K., Torres Beltran, M., Scofield, M., Schwientek, P., Stepanauskas, R., et al. (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell-and meta-genomics. *eLife* 3:e03125. doi: 10.7554/eLife.03125

Roux, S., Krupovic, M., Daly, R. A., Borges, A. L., Nayfach, S., Schulz, F., et al. (2019a). Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol.* 4, 1895–1906. doi: 10.1038/s41564-019-0510-x

Schulz, F., Abergel, C., and Woyke, T. (2022). Giant virus biology and diversity in the era of genome-resolved metagenomics. *Nat. Rev. Microbiol.* 20, 721–736. doi: 10.1038/s41579-022-00754-5

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

Shkoporov, A. N., Clooney, A. G., Sutton, T. D. S., Ryan, F. J., Daly, K. M., Nolan, J. A., et al. (2019). The human gut Virome is highly diverse, stable, and individual specific. *Cell Host Microbe* 26, 527–541.e5. doi: 10.1016/j.chom.2019.09.009

Silva, M., and Baltrusaitis, J. (2021). Destruction of emerging organophosphate contaminants in wastewater using the heterogeneous iron-based photo-Fenton-like process. *J. Hazard. Mater. Lett.* 2:100012. doi: 10.1016/j.hazl.2020.100012

Söding, J. J. B. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21, 951-960. doi: 10.1093/bioinformatics/bti125

Sun, M. M., Yuan, S. J., Xia, R., Ye, M., and Balcázar, J. L. (2023). Underexplored viral auxiliary metabolic genes in soil: diversity and eco-evolutionary significance. *Environ. Microbiol.* 25, 800–810. doi: 10.1111/1462-2920.16329

Suttle, C. A. (2005). Viruses in the sea. Nature 437, 356-361. doi: 10.1038/nature04160

Suttle, C. A. (2007). Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi: 10.1038/nrmicro1750

Tao, Y., Xun, F., Zhao, C., Mao, Z., Li, B., Xing, P., et al. (2023). Improved assembly of metagenome-assembled genomes and viruses in Tibetan saline lake sediment by HiFi metagenomic sequencing. *Microbiol. Spectr.* 11, e03328–e03322. doi: 10.1128/spectrum.03328-22

ter Horst, A. M., Santos-Medellín, C., Sorensen, J. W., Zinke, L. A., Wilson, R. M., Johnston, E. R., et al. (2021). Minnesota peat viromes reveal terrestrial and aquatic niche

partitioning for local and global viral populations. *Microbiome* 9:242. doi: 10.1186/s40168-021-01210-x

Thompson, L. R., Zeng, Q., Kelly, L., Huang, K. H., Singer, A. U., Stubbe, J., et al. (2011). Phage auxiliary metabolic genes and the redirection of *cyanobacterial* host carbon metabolism. *Proc. Natl. Acad. Sci. USA* 108, E757–E764. doi: 10.1073/pnas.1102164108

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

Waite, D. W., Vanwonterghem, I., Rinke, C., Parks, D. H., Zhang, Y., Takai, K., et al. (2018). Addendum: comparative genomic analysis of the class *Epsilonproteobacteria* and proposed reclassification to *Epsilonbacteraeota* (phyl. Nov.). *Front. Microbiol.* 9:772. doi: 10.3389/fmicb.2018.00772

Wang, C. X., and Li, X. (2018). JMT-1: a novel, spherical lytic halotolerant phage isolated from Yuncheng saline lake. *Braz. J. Microbiol.* 49, 262–268. doi: 10.1016/j.bjm.2018.03.004

Wang, C. X., Zhao, A. H., Yu, H. Y., Wang, L. L., and Li, X. (2022). Isolation and characterization of a novel lytic halotolerant phage from Yuncheng Saline Lake. *Indian J. Microbiol.* 62, 249–256. doi: 10.1007/s12088-022-01005-0

Warwick-Dugdale, J., Buchholz, H. H., Allen, M. J., and Temperton, B. J. (2019). Hosthijacking and planktonic piracy: how phages command the microbial high seas. *Virol. J.* 16, 15–13. doi: 10.1186/s12985-019-1120-1

Yang, J., Ma, L., Jiang, H., Wu, G., and Dong, H. J. (2016). Salinity shapes microbial diversity and community structure in surface sediments of the Qinghai-Tibetan Lakes. *Sci. Rep.* 6:25078. doi: 10.1038/srep25078

Yu, M., Zhang, M., Zeng, R., Cheng, R., Zhang, R., Hou, Y., et al. (2024). Diversity and potential host-interactions of viruses inhabiting deep-sea seamount sediments. *Nat. Commun.* 15:3228. doi: 10.1038/s41467-024-47600-1

ZeinEldin, R. A., Ahmed, M. M., Hassanein, W. S., Elshafey, N., Sofy, A. R., Hamedo, H. A., et al. (2023). Diversity and distribution characteristics of viruses from soda lakes. *Genes* 14:323. doi: 10.3390/genes14020323

Zeng, F., Zhu, Y., Zhang, D., Zhao, Z., Li, Q., Ma, P., et al. (2022). Metagenomic analysis of the soil microbial composition and salt tolerance mechanism in Yuncheng salt Lake, Shanxi Province. *Front. Microbiol.* 13:1004556. doi: 10.3389/fmicb.2022. 1004556

Zhang, X., Gao, X., Li, C., Luo, X., and Wang, Y. (2019). Fluoride contributes to the shaping of microbial community in high fluoride groundwater in Qiji County, Yuncheng City, China. *Sci. Rep.* 9:14488. doi: 10.1038/s41598-019-50914-6

Zhang, X. Y., Zhang, R., Wang, J. C., Zhang, T., and Du, Z. J. (2020). Wenzhouxiangella limi sp. nov., isolated from a salt lake. Int. J. Syst. Evol. Microbiol. 70, 4610–4615. doi: 10.1099/ijsem.0.004320

Zheng, M., and Liu, X. (2009). Hydrochemistry of salt lakes of the Qinghai-Tibet plateau, China. Aquat. Geochem. 15, 293-320. doi: 10.1007/s10498-008-9055-y

Zhou, Y. L., Mara, P., Vik, D., Edgcomb, V. P., Sullivan, M. B., and Wang, Y. (2022). Ecogenomics reveals viral communities across the challenger deep oceanic trench. *Commun. Biol.* 5:1055. doi: 10.1038/s42003-022-04027-y

Zhou, H., Tang, J., Guo, J., Dai, Y., Li, G., and Yan, B. (2019). Integrated system of comprehensive utilizing the concentrated brine of Yuncheng salt-lake basing on salt-forming diagram. *Chin. J. Chem. Eng.* 27, 182–190. doi: 10.1016/j.cjche. 2018.03.008

Zimmerman, A. E., Howard-Varona, C., Needham, D. M., John, S. G., Worden, A. Z., Sullivan, M. B., et al. (2020). Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol.* 18, 21–34. doi: 10.1038/s41579-019-0270-x