Environmental Science and Ecotechnology 30 (2026) 100683

Contents lists available at ScienceDirect

Environmental Science and Ecotechnology

journal homepage: www.journals.elsevier.com/environmental-science-and-ecotechnology/

Original Research

Hi-C sequencing deciphers phage and plasmid host networks in wastewater biofilms

Dou Wang [a], Xiaoqing Xu [a], Lei Liu [a], Chunxiao Wang [a], Yu Deng [a,b], Martin F. Polz [c,d], Tong Zhang [a,e,f,g,h,*]

[a] Environmental Microbiome Engineering and Biotechnology Laboratory, Center for Environmental Engineering Research, Department of Civil Engineering, The University of Hong Kong, Hong Kong Special Administrative Region of China
[b] Faculty of Dentistry, The University of Hong Kong, Hong Kong Special Administrative Region of China
[c] Division of Microbial Ecology, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria
[d] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[e] School of Public Health, The University of Hong Kong, Hong Kong Special Administrative Region of China
[f] Macau Institute for Applied Research in Medicine and Health, Macau University of Science and Technology, Macau Special Administrative Region of China
[g] State Key Laboratory of Marine Pollution, City University of Hong Kong, Hong Kong Special Administrative Region of China
[h] Shenzhen Institute of Research and Innovation, The University of Hong Kong, Shenzhen, China

ARTICLE INFO

Article history:
Received 30 October 2025
Received in revised form
5 March 2026
Accepted 5 March 2026

Keywords:
Phage
Plasmid
Biofilm
Assembly
Interactions

ABSTRACT

Mobile genetic elements (MGEs) such as bacteriophages and plasmids profoundly shape microbial community structure and drive horizontal gene transfer across ecosystems. Wastewater treatment systems, with their high cell densities, steep physicochemical gradients and close cell-to-cell contact, act as hotspots for MGE proliferation and exchange, yet the *in situ* assembly dynamics and host interaction networks of these elements have remained largely unresolved because conventional methods fail to establish direct MGE–host linkages in complex matrices. Here we show that an integrated framework combining metagenomics, metatranscriptomics, metaviromics, and Hi-C proximity ligation sequencing enables the efficient elucidation of DNA phage and plasmid assembly dynamics alongside their host interaction networks in biofilms. We reconstructed 17,672 viral operational taxonomic units and 11,454 high-confidence non-redundant plasmids, and established 529 phage–host and 5739 plasmid–host associations that link up to 52 % of phages to 56 % of prokaryotes and 70 % of plasmids to 91 % of prokaryotes, respectively. Hi-C substantially expanded and refined these networks, revealing taxon-specific and multi-host patterns. Host community composition and biofilm architecture emerge as primary drivers of MGE occurrence and abundance along the reactor flow path. Expression of auxiliary metabolic genes, antibiotic resistance genes and virulence factors carried by these MGEs demonstrates their active roles in modulating biogeochemical cycles and maintaining ecosystem stability. These findings establish a scalable, cultivation-independent framework for deciphering MGE–host networks in complex microbial ecosystems, and underscore the power of Hi-C sequencing to transform our mechanistic understanding of gene flow, resistome dissemination, and ecological resilience in engineered and natural microbiomes.

1. Introduction

Bacteriophages (phages) and plasmids are mobile genetic elements (MGEs) that play key roles in shaping microbial community composition and mediating horizontal gene transfer (HGT) [1–3]. Phages modulate microbial dynamics via lytic–lysogenic cycles and influence biogeochemical processes by altering host metabolic pathways [4–6]. Plasmids serve as key agents for disseminating

* Corresponding author. Environmental Microbiome Engineering and Biotechnology Laboratory, Center for Environmental Engineering Research, Department of Civil Engineering, The University of Hong Kong, Hong Kong Special Administrative Region of China.
E-mail address: zhangt@hku.hk (T. Zhang).

https://doi.org/10.1016/j.ese.2026.100683

adaptive traits, including antibiotic resistance genes (ARGs), virulence factors (VFs), and metal resistance determinants [7,8]. Notably, plasmids are recognized as critical vectors for HGT and the spread of ARGs, with over 70% of ARGs being plasmid-borne [9]. Understanding the ecological impacts of MGEs fundamentally depends on resolving their host associations and interaction networks, yet in complex microbial communities, such as wastewater treatment systems, these relationships remain poorly mapped and mechanistically underexplored.

Wastewater treatment systems are recognized as hotspots for MGE proliferation and exchange, hosting MGE diversity far exceeding that in most natural ecosystems [10,11]. Biofilms, with their high cellular density, structural heterogeneity, and pronounced physicochemical gradients, offer a compelling model system for investigating MGE–host interactions [12,13]. These structured communities facilitate close cell-to-cell contact, promote plasmid conjugation, and influence phage–host encounter rates, thereby fostering intricate networks of genetic exchange and coevolution [14]. However, precisely how phage and plasmid communities assemble in biofilms, how they distribute across host taxa, and how their interactions affect community function and stability have yet to be determined.

A major bottleneck lies in the lack of methods capable of simultaneously and accurately identifying hosts for phages and plasmids *in situ*, especially within multispecies biofilm matrices. Culture-dependent approaches are constrained by low microbial culturability and scalability, especially in complex ecosystems. Molecular techniques, such as fluorescence *in situ* hybridization and targeted polymerase chain reaction, require prior sequence knowledge and offer limited scalability. Computational prediction methods, while cost-effective, have accuracy challenges due to data quality biases and algorithmic assumptions: clustered regularly interspaced short palindromic repeats (CRISPR) spacer matching applies only to 40–50% of bacteria possessing CRISPR-Cas systems [15–17], and machine learning models suffer from "black-box" interpretability issues [18,19]. Although emerging single-cell sequencing can directly link hosts to MGEs, it is technically demanding, low-throughput, and difficult to apply to dense matrix-embedded biofilms [20,21]. Recently, Hi-C sequencing has emerged as a powerful technique for capturing chromatin proximity within intact microbial communities, enabling the direct physical linkage of MGEs to their host chromosomes without prior cultivation [22–24]. The utility of Hi-C has been demonstrated across diverse ecosystems, including wastewater, soil, and marine environments, where it has successfully resolved intricate MGE–host relationships [23,25,26]. When integrated with metagenomics and metatranscriptomics, Hi-C offers a transformative path for reconstructing MGE–host networks while preserving the ecological context.

In the present study, we leveraged the complementary strengths of diverse methods by employing an integrated multi-omics and Hi-C framework to systematically resolve the assembly dynamics and host interaction networks of phages and plasmids within a model biofilm ecosystem. We successfully reconstructed 17,672 viral operational taxonomic units (vOTUs) and 11,454 plasmid sequences, highlighting their extensive diversity and distribution within this system. By combining Hi-C and *in silico* prediction, we identified 529 phage–host and 5739 plasmid–host associations, revealing tight interactions between these MGEs and prokaryotes. This comprehensive investigation deepens our understanding of the assembly dynamics and ecological functions of phages and plasmids within biofilm systems and lays the groundwork for future research to harness these interactions to optimize biofilm functionality and mitigate the spread of ARGs through HGT.

2. Materials and methods

2.1. Sample collection, sequencing, and Hi-C library preparation

Biofilm sampling was conducted on March 19, 2020, at the Ma Wan Sewage Treatment Plant in Hong Kong, which processes approximately 7600 m^3 of wastewater daily. Samples were collected from nine predetermined points (designated A1–A4 and B1–B5) along two concatenated rotating biological contactor (RBC) trains (Stages A and B) covering the complete flow path of the system, as outlined in our previous article [27]. At each designated sampling point along the rotary disk, biofilms were harvested from several circumferential subsites using a sterilized metal brush. The collected biofilm samples from multiple subsites belonging to the same sampling point were pooled into a 50 mL centrifuge tube and homogenized thoroughly to form a composite sample representative of that sampling point. This composite sample was immediately flash-frozen in liquid nitrogen, transported to the laboratory, and stored at −80 °C until DNA extraction and subsequent sequencing analysis. The same standardized procedure was consistently applied across all sampling points to ensure comparability between samples. In addition, water samples were collected at each sampling point, and the chemical parameters for these locations were measured. The detailed results for these measurements, as well as information on nucleotide extraction, shotgun metagenomic sequencing, long-read sequencing, and metatranscriptomic sequencing, can be found in our previous article [27].

For Hi-C library preparation, the collected biofilm samples were first fixed with 1% formaldehyde, and excess glycine was added to quench the formaldehyde. The samples were then recovered by centrifugation, rinsed with phosphate-buffered saline, recentrifuged, and stored at −80 °C. Subsequently, the fixed samples were delivered to Phase Genomics (Seattle, WA 98109, USA) for treatment and sequencing. The detailed library preparation process is described in Supplementary Text S1.

2.2. Prokaryotic-genome reconstruction

In total, we obtained 97 gigabases (Gb) of short reads and 96 Gb of long reads from the nine biofilm samples. Metagenomic paired-end reads were processed using fastp (v0.20.1) [28] with the default settings before analysis. For genome reconstruction, the iterative, haplotype-resolved, hierarchical-clustering-based hybrid assembly approach was employed in the present study [27,29]. We clustered the metagenome-assembled genomes (MAGs) across all locations at 99% average nucleotide identity (ANI) using dRep (v2.6.2) [30] to generate a nonredundant genome set. The quality of these MAGs was assessed with CheckM (v1.1.3) [31], and only those that met the thresholds of $\geq$50% completeness and $\leq$10% contamination were retained as representative MAGs (rMAGs) for all downstream analyses, including host identification. The detailed genome classification, annotation, and metabolic prediction processes are described in Supplementary Text S2 and S3.

2.3. Plasmid and ARG annotation

Short reads from the nine biofilms were initially assembled using metaSPAdes (v3.15.5) [32] to generate contigs. To improve assembly continuity, long reads were subsequently integrated using OPERA-MS (v0.9.0) [33] to produce longer contigs, followed by error correction with Pilon (v1.22) [34] using short reads. Contigs $\geq$1 kb were retained for downstream analysis. Before plasmid identification, the contigs generated from each sample

and the reconstructed MAGs were pooled and clustered (95% ANI, 85% alignment fraction) to obtain composite contigs. Subsequently, geNomad (v1.6.1) [2] was employed to detect candidate plasmids with default settings. To ensure the specificity of plasmid detection, candidate plasmid sequences were validated using PlasFlow (v1.1) [35], with sequences annotated as chromosomal origins excluded from subsequent analyses. For ARG annotation, open reading frames predicted by Prodigal (v2.6.3) [36] were aligned against the structured-ARG database [37] using DIAMOND BLASTP with the –id 80 and –query-cover 70 cutoffs. Using DIAMOND BLASTP with the same cutoff applied to the ARG annotation, we annotated VFs by searching against the full protein dataset of the VF database [38].

2.4. Viral sample collection, enrichment, and DNA processing

To establish a comprehensive phage catalog, we systematically collected samples from both the biofilm and liquor phases of the system. The biofilm samples collected from Stage A (A1–A4) and Stage B (B1–B5) were pooled to create two composite biofilm samples. Concurrently, to maximize the recovery of free phage particles, we collected 5 L of liquor samples from Stages A and B along the flow path. The detailed protocols for sample pretreatment and phage enrichment are described in our previous article [10,27]. Briefly, the samples underwent sonication to detach the phage particles, followed by sequential filtration through 0.45 μm and 0.22 μm membranes. The phage particles were then precipitated using polyethylene glycol, and the enriched samples were used for downstream DNA extraction. We extracted phage DNA from the three enriched samples using a commercial phage DNA isolation kit (Norgen Biotek Corp., Canada) and stored it at −80 °C until further processing. Metagenomic sequencing was performed on the Illumina NovaSeq PE150 platform (Novogene, Beijing, China), yielding 150 bp paired-end reads. In total, approximately 30 Gb of high-quality short-read data were obtained for downstream analysis.

2.5. Phage identification, classification, and annotation

The three metaviromes were *de novo* co-assembled using metaSPAdes (v3.13.0) [32] to generate candidate phage contigs. We excluded contigs smaller than 5 kb from the downstream analyses. To identify phage sequences, we applied two different approaches: VirSorter2 (v2.1) [39] and a modified virus DNA detection pipeline utilizing virus-specific Hidden Markov Models [40]. The phage contigs identified by both approaches were consolidated to obtain a comprehensive collection of phage-contig candidates. These composite phage contigs were then manually curated according to the screening criteria proposed by the Sullivan Lab (dx.doi.org/10.17504/protocols.io.bwm5pc86) to obtain a reliable set of trusted phage contigs. Subsequently, we performed UCLUST-like clustering on the phage set to yield nonredundant representative vOTUs, using the Minimum Information about an Uncultivated Virus Genome-recommended parameters of 95% ANI over an 85% alignment fraction (relative to the shorter sequence) [41]. The resulting vOTU set was used for all subsequent downstream analyses.

Before conducting downstream analyses, potential host contamination in the vOTUs was removed with CheckV (v0.5.1) [42]. Taxonomy classification of the vOTUs was conducted using gene-sharing network clustering implemented in vConTACT2 (v0.9.19) [43] with the ProkaryoticViralRefSeq201-Merged database, complemented by PhaGCN (v2.3) [44] based on the latest International Committee on Taxonomy of Viruses database. Gene annotation and prediction of potential auxiliary metabolic genes (AMGs) within the vOTUs were carried out using DRAM-v (v1.2.4) [45]. Prior to annotation, we processed the vOTUs with VirSorter2 (v2.2.2) [39] to generate the necessary inputs for DRAM-v, including a VirSorter2-predicted vMAGs FASTA file and a VIRSorter_affi-contigs.tab file. Annotation was performed based on the default database, and the results were distilled using the DRAM-v distill function. To further validate the potential functional roles of AMGs encoded by vOTUs in the ecosystem, the secondary and tertiary structures of AMGs associated with vOTUs linked to specific hosts were predicted using Phyre2 (v2.0) [46]. Additionally, the lifestyle of the vOTUs (lytic or lysogenic) was determined using BACPHLIP [47] and PhaTYP [48]. We prioritized predictions from PhaTYP and used BACPHLIP results only when PhaTYP provided no prediction.

2.6. Prediction of phage– and plasmid–host association

We applied Hi-C sequencing and three other *in silico* prediction methods (CRISPR-spacer, homology, and transfer RNA [tRNA] matches) to identify putative phage–host associations. The ProxiMeta platform, developed by Phase Genomics (Seattle, WA 98109, USA), was used to assign hosts based on Hi-C data (see Supplementary Text S4 for details). CRISPR spacers were identified in the recovered rMAGs using the CRISPR Recognition Tool [49]. Then, we searched the extracted spacers against the vOTUs using BLASTn (BLASTn-short, ≥97% identity, ≥90% coverage, and ≤1 mismatch) [10]. The contigs of the recovered rMAGs were compared with the vOTUs using BLASTn (≥90% identity, matches ≥2500 bp) to find the homology sequences [50–52]. The tRNAs in the vOTUs were identified using tRNAscan-SE (v2.0.9) [53] with the general tRNA model (option -G, cytosolic tRNAs from all three domains included) with the default parameters. We then queried the identified tRNAs against the rMAGs using BLASTn with 100% query coverage and 100% sequence identity [54]. The phage–host pairs identified by each method were then combined into a composite collection of interactions. Plasmid–host identification was performed using the Hi-C method and CRISPR spacer matching [1], following the criteria established for phage–host identification.

2.7. Metatranscriptomic analysis

We applied SortMeRNA (v4.0.0) [55] to filter noncoding RNA sequences from metatranscriptomic data using comprehensive ribosomal RNA databases containing bacterial, archaeal, and eukaryotic sequences. Gene abundance and transcriptional activity associated with phage and plasmid elements were quantified using two metrics: genes per million [56] and transcripts per million, calculated via the RSEM pipeline [57]. For phage-associated genes, all predicted phage genes were aggregated for integrated expression analysis. To investigate plasmid-borne functional and accessory genes, including mobilization/conjugation-related genes and VFs, all plasmid-derived genes were consolidated for comparative profiling. Given the predominant chromosomal localization of ARGs observed in the present study, all ARG-carrying contigs were pooled for coverage calculation and expression quantification.

2.8. Compositional analysis of prokaryotes and MGEs

For the prokaryotic community, we employed the relative abundances of the rMAGs derived from biofilm metagenome alignments using CoverM (v0.4.0) [58] for principal coordinate analysis (PCoA). For the analysis of the MGE community distribution pattern, we mapped bulk biofilm metagenomic reads to the vOTUs/plasmids using CoverM (v0.4.0) [58] in "-mean" mode to

calculate the coverage (divided by the contig length). The parameters were set to 90% read-percent identity and 80% read-aligned percent. To determine the presence of a vOTU/plasmid in the biofilm samples, we implemented a tiered coverage-threshold strategy tailored to the analytical purpose. For broad ecological patterns (e.g., α- and β-diversity), we considered a vOTU or plasmid present if aligned reads covered ≥10% of its length, consistent with thresholds used in the Earth Virome Project [54] to retain sensitivity for detecting low-abundance viruses in non-enriched metagenomes. For higher-confidence downstream analyses, such as phage/plasmid–host interaction and co-occurrence inference, a more stringent threshold of ≥50% coverage was applied to reduce false positives. We further normalized coverage values by the metagenomic dataset volume, defined as coverage per Gb of short-read metagenome [51]. The resulting coverage matrix was utilized for alpha diversity calculation and PCoA using the R package [59]. Furthermore, we conducted a Procrustes analysis to explore the associations between prokaryotes and phages/plasmids with the system.

3. Results

3.1. Harboring of diverse and novel phages by the biofilm system

We co-assembled three metaviromic datasets generated from biofilm and liquor samples using phage particle enrichment and screened the assemblies for phage identification. VirSorter2 and a modified pipeline identified 17,509 and 12,542 putative phage contigs (≥5 kb), respectively. After consolidating these results, we manually curated and clustered the contigs to generate 17,672 high-confidence vOTUs for downstream analysis. Details of the final vOTU set are provided in Supplementary Table S1.

The final vOTU sets could recruit 44.6–57.7% of the three metaviromic samples (Supplementary Fig. S1), whereas only 0.2–0.5% of the nine metagenomic reads could be aligned. This demonstrated that our approach efficiently enriched the phage particles from the biofilm samples. Combining the results from both lifestyle prediction tools, 4845 (27%) vOTUs were inferred as potential lysogenic, while the remaining were predicted as lytic (Fig. 1a and Supplementary Table S1). The CheckV results revealed that only 362 and 415 vOTUs were estimated to be complete and high-quality, respectively (Fig. 1a and Supplementary Table S1), with 442 recognized as prophages. Gene-sharing network clustering yielded 2101 viral clusters (VCs) with two or more members (Supplementary Fig. S2a and Table S2); these clusters contained 6590 vOTUs reconstructed in the present study, indicating high phage diversity. Of these clusters, only 64 VCs could be classified into known viral taxa based on the NCBI RefSeq viral genome database, implying the novelty of the reconstructed vOTUs (Supplementary Fig. S2b). According to the classification, 63 VCs were Caudoviricetes, and only a limited number could be assigned to the family level. PhaGCN2 generated 13,333 predictions at the class level, accounting for approximately 75% of the vOTUs (Fig. 1a). Similar to the classification result of vConTACT2, over 96% of the annotated vOTUs were assigned to the class of Caudoviricetes. vOTUs affiliated to the families of Steigviridae, Autographiviridae, Schitoviridae, Herelleviridae, Zobellviridae, and Suoliviridae accounted for ~43% of the taxonomically classified vOTUs (Fig. 1a and Supplementary Table S3).

We analyzed bulk biofilm metagenomes to evaluate the phage distribution and diversity across the biofilm system. Approximately 0.2–0.5% of the nine metagenomic reads were recruited by the reconstructed vOTUs, with prophage accounting for only 0.003–0.01% of the mapped reads. To compare the phage community compositions across the reactor, we used a coverage matrix derived from these alignments, which included only the vOTUs present in at least two samples (Supplementary Table S4). The PCoA results revealed distinct segregation between the Stage A and B phage communities (Fig. 1b and Supplementary Fig. S3), with adjacent sampling points within each stage exhibiting greater compositional similarity. Phage α-diversity, as measured by Shannon indices, was elevated in both the influent (front) and effluent (rear) regions of the reactor (Fig. 1c and Supplementary Fig. S4). Notably, a higher total coverage of phages was observed at Stage A (from A2 to A4) (Fig. 1d), although the ratios of virulent and temperate phages remained stable across all locations. The lower Shannon indices observed in the high-coverage locations suggest a predominance of highly abundant phage populations.

3.2. Highly linked phages and prokaryotes

To ensure high accuracy in predicting phage–host relationships, we only used rMAGs recovered from the biofilm system, rather than incorporating external database genomes. Metagenomic sequencing, assembly, and binning across all locations yielded a dereplicated set of 525 prokaryotic rMAGs (≥50% completeness and ≤10% contamination). The genomic characteristics and assembly statistics for all rMAGs are comprehensively documented in Supplementary Table S5, with the phylogenetic placement visualized in Supplementary Fig. S5. Subsequently, four approaches were integrated to identify phage–host associations: spacer search, homology match, tRNA search, and phage–host DNA crosslinking (Hi-C analysis) (Fig. 2a and Supplementary Table S6). The hybrid assembly strategy enabled comprehensive detection of CRISPR arrays and tRNA sequences (Supplementary Table S5), potentially enabling resolution of phage–host associations. Specifically, the spacer- and homology matches-based searches resulted in 87 and 48 phage–host pairs, respectively. Additionally, 12 phage–host pairs were identified by comparing tRNA genes. Notably, Hi-C analysis substantially expanded the interaction network, identifying 177 rMAGs connected to 400 vOTUs (429 phage–host pairs), nearly doubling the associations detected by other methods combined. Following dereplication, the integration of all approaches yielded 529 phage–host interactions (487 vOTUs–210 rMAGs), with 41 associations (7.8%) being cross-validated by ≥ 2 independent methods.

Approximately 30–52% of the vOTUs (coverage-weighted) demonstrated specific host associations, corresponding to 38–56% of the reconstructed rMAGs (coverage-weighted) across all biofilm samples (Fig. 2b). The phage interactions displayed a clear taxon-specific pattern (Fig. 2b and Supplementary Fig. S6 and S7). Pseudomonadota and Bacteroidota are the most prevalent hosts at the phylum level (Fig. 2b). This observation aligns with the overall dominance of these prokaryotic communities across all sampled locations (Supplementary Fig. S5c). Notably, Myxococcota-affiliated rMAGs ABC2.88 and BC3.13 exhibited polyvalent interactions with multiple vOTUs, while their phylum displayed significantly elevated phage-to-host ratios compared to other taxa (Supplementary Fig. S8). Spatial analysis revealed strong phage–host co-occurrence patterns along the reactor's flow path (Fig. 2b). Phages associated with Nitrospirota, Acidobacteriota, Thermoproteota, and Bdellovibrionota showed strict spatial coupling with their hosts, with phage abundance increasing at the sampling locations where hosts were present. For instance, the high abundance of comammox *Nitrospira* and ammonia-oxidizing archaea at Stage B coincided with their respective phage populations. Procrustes analysis confirmed significant community-level coordination between prokaryotic and phage assemblages ($M^2 = 0.051$, $p = 0.001$) (Fig. 2c and Supplementary Fig. S9), indicating a tight connection between these biological compartments.

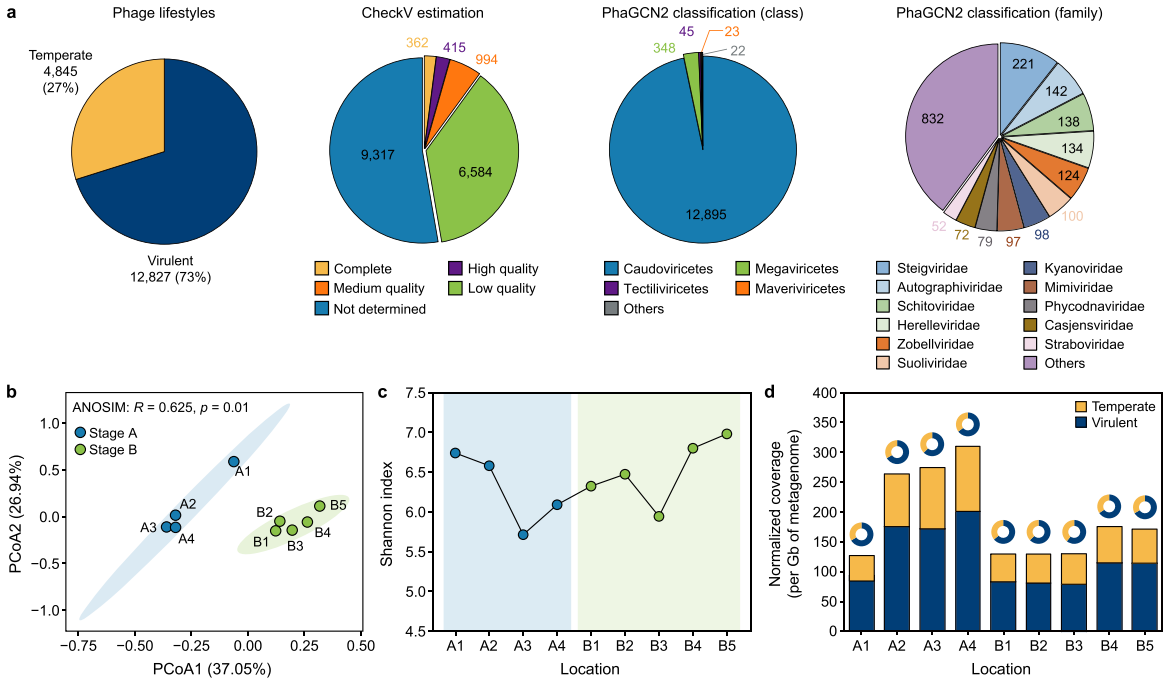

Fig. 1. Phage catalog establishment and profiles in biofilm metagenomes. **a,** Overview of the reconstructed viral operational taxonomic units (vOTUs), including quality assessment using CheckV, taxonomic classification via PhaGCN2, and lifestyle prediction. See Materials and Methods section for details. **b,** Principal coordinate analysis (PCoA) of phage communities based on Bray–Curtis dissimilarities (derived from normalized vOTU coverage), highlighting compositional differences between Stage A and Stage B. Differences were assessed using analysis of similarities (ANOSIM). Ellipses represent 95% confidence intervals for phage communities in each stage. Only vOTUs detected in at least two samples were included. **c,** Shannon diversity indices of phage communities derived from the normalized vOTUs coverage matrix. **d,** Normalized total vOTU coverage across sampling locations. Bar plots show the coverage of temperate (yellow) and virulent (deep blue) vOTUs at each location. Doughnut charts represent the corresponding proportions of temperate and virulent vOTUs per location.

3.3. Phages' mediation of biogeochemical cycling by modulating host turnover and metabolism

Phages have been documented to exert modulatory effects on biogeochemical cycles, employing strategies that include host lysis, metabolic reprogramming, and the expression of AMGs. To validate their potential role within this biofilm system, we first characterized the metabolic profiles of phage-associated prokaryotic hosts. Genomic analysis revealed that these identified prokaryotic hosts constitute a substantial proportion of functional microbiota, with a notable enrichment in nitrogen transformation pathways (Supplementary Fig. S10). This suggests that phages may influence biogeochemical cycling by targeting key microbial populations involved in nutrient processing. Furthermore, we identified 3213 AMGs distributed across 2007 vOTUs, of which 1146 genes were successfully assigned to specific category modules (Supplementary Table S7). Functional categorization demonstrated predominant AMG involvement in three core metabolic processes: carbon utilization, miscellaneous, and organic-nitrogen metabolism (Fig. 3). Specifically, 412 AMGs from 356 vOTUs encode enzymes for carbon utilization. Among these, 77 genes encode glycoside hydrolases—enzymes that may depolymerize complex carbohydrates, such as starch (via the backbone cleavage), xylan, and rhamnose oligomers, during phage infection (Supplementary Table S7). We also identified AMGs involved in C1 metabolism (e.g., reductive citrate and pentose phosphate cycles), amino acid biosynthesis, hydrocarbon degradation, and various transport systems. This diverse AMG repertoire implies broad potential for phage-mediated metabolic augmentation within the biofilm.

To move beyond genetic potential and assess *in situ* activity, we analyzed AMG transcription using metatranscriptomic data. To enhance the reliability of the inferred biogeochemical functions, genes associated with glycosyl transferases, nucleotide metabolism (a core component of the information system), organic-nitrogen processing, and ribosomal proteins were excluded from the analysis [4,60,61] because these are commonly linked to viral replication, host attachment, and cellular entry [60,61]. The results showed that phage-encoded glycoside hydrolases and polysaccharide lyases were ubiquitously expressed across all sampling locations (Supplementary Table S8), implying that phages may optimize the host's carbohydrate metabolism pathway during infection. Additionally, a relatively highly expressed gene involved in hydrocarbon degradation was observed (Fig. 3), especially at site A1. Notably, we detected the expression of carbohydrate esterases encoded by vOTUs linked to prokaryotic hosts, further hinting at their potential in facilitating host metabolism. Moreover, the tertiary structures of AMGs carried by phages that can be linked to specific hosts also support the possibility that these AMGs are functional in the environment (Supplementary Fig. S11).

3.4. Dynamic pattern of plasmidome and resistome across the studied biofilm system

A total of 12,061 sequences derived from clustered contigs were annotated as putative plasmids, of which 11,454 high-confidence plasmid sequences were ultimately retained after rigorous decontamination to remove chromosome-like sequences (Supplementary Table S9). The curated plasmidome exhibited a mean length of 11,693 bp, with 32 plasmids confirmed as circular through terminal repeat analysis. Functional annotation revealed hallmark plasmid maintenance genes in 2254 (19.7%) plasmids and conjugation-associated genes in 1513 (13.2%) plasmids (Supplementary Table S9). Notably, conjugation-capable plasmids

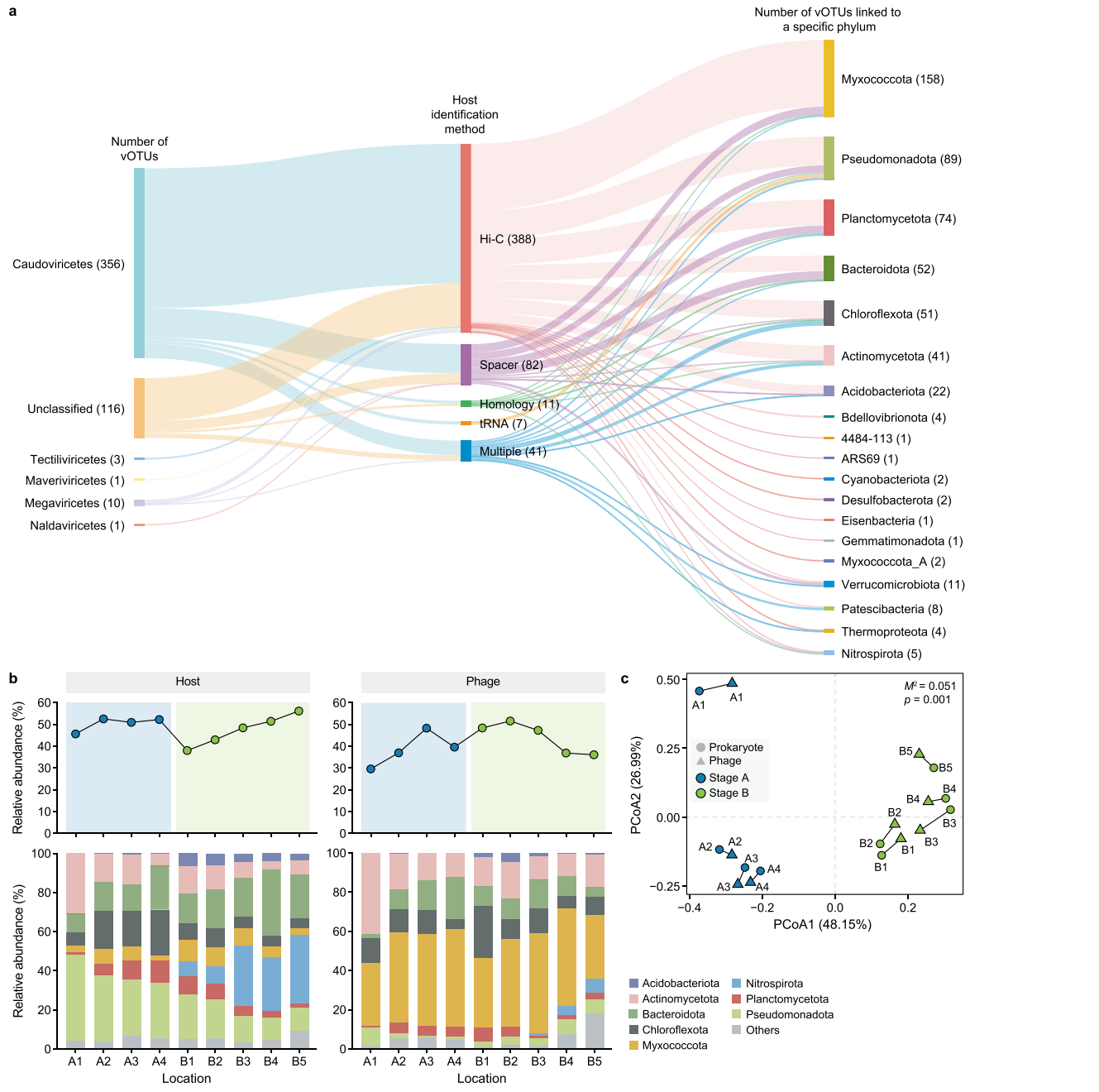

Fig. 2. Phage–host interaction network and community dynamics. **a,** Sankey diagram illustrating phage–host associations predicted using four different methods—totaling 529 interactions, connecting 487 vOTUs and 210 prokaryotic hosts. Left, number of viral operational taxonomic units (vOTUs) assigned to each phage taxon for which a prokaryotic host has been successfully identified. Center, number of phage–host links detected by each method. Right, number of vOTUs linked to each host phylum. Numbers are shown in parentheses. **b,** Phylum-level relative abundance of host organisms (left) and their associated phages (right) across sampling locations. **c,** Procrustes analysis comparing structural congruence between prokaryotic and phage communities. Prokaryotic communities were defined by representative metagenome-assembled genomes (rMAGs); only rMAGs and vOTUs detected in at least two samples were included.

accounted for 13.7–22.1% of the total plasmid sequence coverage across the sampling points, despite representing only 13.2% of the total plasmid count. Among the mobilization (MOB) families, MOBP1 dominated the repertoire of conjugation genes, accounting for 51% of all annotated MOB loci.

We profiled plasmid distribution patterns using a read-recruitment matrix constructed from metagenomic alignments (Supplementary Table S10). The final plasmid collection recruited 1.9–3.6% of the metagenomic reads across all samples. Intriguingly, plasmid coverage showed a clear linear decline at Stage A, followed by an increase at Stage B (Fig. 4a). This was inconsistent with the trend in plasmid diversity (Shannon index) across both stages (Fig. 4b), suggesting selective displacement of dominant plasmids rather than community-wide restructuring. PCoA demonstrated a clear separation in plasmidome profiles between Stages A and B (analysis of similarities, $R = 0.694$, $p = 0.006$) (Fig. 4c). Procrustes analysis revealed coordinated dynamics between plasmid communities and prokaryotic microbiome shifts ($M^2 = 0.024$, $p = 0.001$) (Fig. 4d), demonstrating consistent plasmid–host associations.

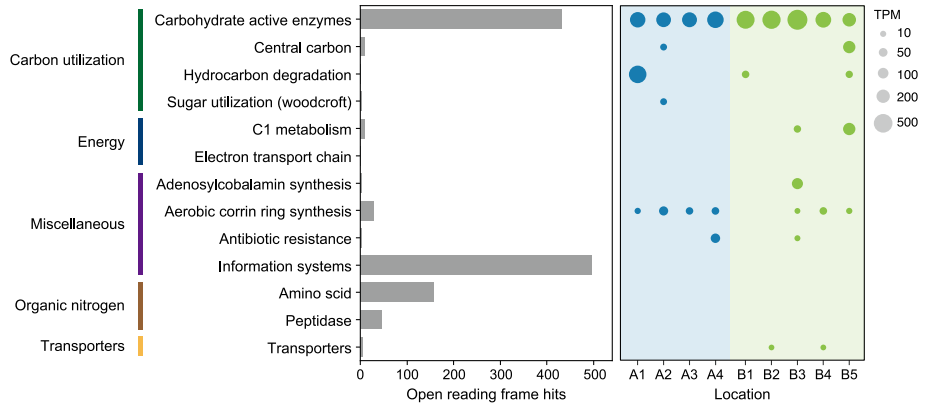

Fig. 3. Functional annotation and expression profiles of auxiliary metabolic genes (AMGs) identified from reconstructed viral operational taxonomic units (vOTUs). The categories are classified based on the DRAM-v annotation result. TPM, transcripts per million.

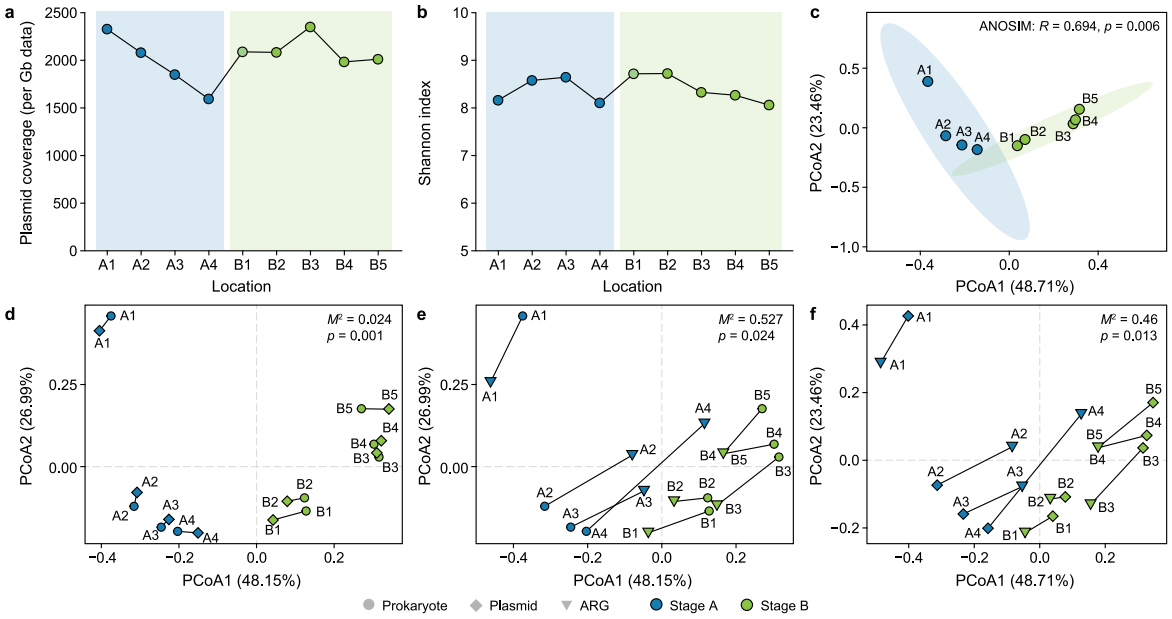

Fig. 4. Plasmidome and resistome dynamics in biofilm metagenomes. **a,** Normalized total plasmid coverage across sampling locations. **b,** Shannon diversity indices of plasmid communities, calculated from the normalized plasmid coverage matrix. **c,** Principal coordinate analysis (PCoA) of plasmid composition based on Bray–Curtis dissimilarities (derived from normalized plasmid coverage), illustrating distinctions between Stage A and Stage B. Only plasmids detected in at least two samples were included for this analysis. Analysis of similarities (ANOSIM) was used to assess differences in the plasmid community between stages. Ellipses represent 95% confidence intervals for plasmid communities in each stage. **d–f,** Procrustes analysis of structural congruence between prokaryotic and plasmid communities (**d**), prokaryotic and ARG collections (**e**), and plasmid and ARG collections (**f**). Only rMAGs and/or plasmids detected in at least two samples were included in the respective analyses.

Comprehensive resistome profiling via the ARGs-OAP [37] revealed conserved ARG composition along the flow path, despite quantitative abundance fluctuations (Supplementary Fig. S12). Peak ARG load occurred at the initial sampling point (A1), with subsequent sites showing reduced but comparable coverage levels. The PCoA results based on normalized ARG subtype abundance confirmed the distinctiveness of A1's resistome profile (analysis of similarities, $R = 0.444$, $p = 0.021$) (Supplementary Fig. S13). While moderate concordance existed between ARG distribution and prokaryotic-community structure (Fig. 4e), this association was significantly weaker than the observed plasmid–prokaryote connection. The investigation of plasmid correlation yielded similar results (Fig. 4f). This attenuated connectivity suggests that ARG dissemination may be less constrained by host specificity, potentially facilitated by HGT. Furthermore, contig-level analysis identified 191 ARGs, 38 of which were located on plasmids (Supplementary Table S11).

3.5. Extensive plasmid–host associations

Integrated analysis through three complementary approaches—Hi-C capture, CRISPR spacer matching, and co-binning from rMAGs—revealed extensive plasmid–host associations. Hi-C analysis provided the highest resolution, detecting 5368 high-confidence linkages connecting 4738 plasmids with 418 rMAGs. Spacer matching and MAG co-binning identified 46 and 351 additional associations, respectively, with 23 spacer-derived pairs (50%) showing Hi-C validation (Supplementary Table S12). The combined network comprised 5739 unique interactions linking 5096 plasmids to 428 rMAGs (Fig. 5a), demonstrating both methodological complementarity and plasmid multi-host potential. Pseudomonadota emerged as the dominant plasmid host

phylum, accounting for 68.3% of all associations (3920/5739). Within this phylum, 160 host populations (37.4% of all identified hosts) were associated with 3572 plasmids, averaging ~22 plasmids per host population, a value significantly higher than that observed in other phyla. Notably, 591 plasmids demonstrated multi-host interactions (≥2 hosts), including 48 plasmids with broad connectivity (≥3 hosts). Among these plasmids, 21 exhibited a cross-phylum host range, while 47 spanned multiple genera, highlighting their capacity to interact with taxonomically distinct hosts. Host network analysis identified 38 potential plasmid-mediated ARG dissemination routes across 27 plasmids. The most frequently transmitted resistance determinants were *aph(3″)-Ib* ($n = 6$), *aph(6)-Id* ($n = 6$), and *tet(X)* ($n = 5$). Notably, 29.6% (8/27) of these ARG-carrying plasmids demonstrated multi-host transmission, suggesting their role as vectors for resistance dissemination.

Quantitative linkage analysis revealed that 54–70% (coverage-weighted) of the plasmid populations exhibited detectable host associations through our multi-method framework, covering 78–91% (coverage-weighted) of the biofilm-associated prokaryotes (Fig. 5b). Pseudomonadota, Bacteroidota, and Nitrospirota as dominant plasmid hosts collectively accounted for 48–62% of the identified host populations (Fig. 5b). Notably, by leveraging established plasmid–host linkages, we investigated plasmid carriage at the single-cell level using plasmid-to-host coverage ratios as a metric. Our analysis revealed significant phylum-level variation in these ratios, with Pseudomonadota exhibiting the highest plasmid carriage capacity (Supplementary Fig. S14). This disparity highlights fundamental differences in plasmid maintenance strategies across taxa. Furthermore, the declining trend observed in the Pseudomonadota ratio aligned with the overall reduction in plasmid coverage at Stage A. The

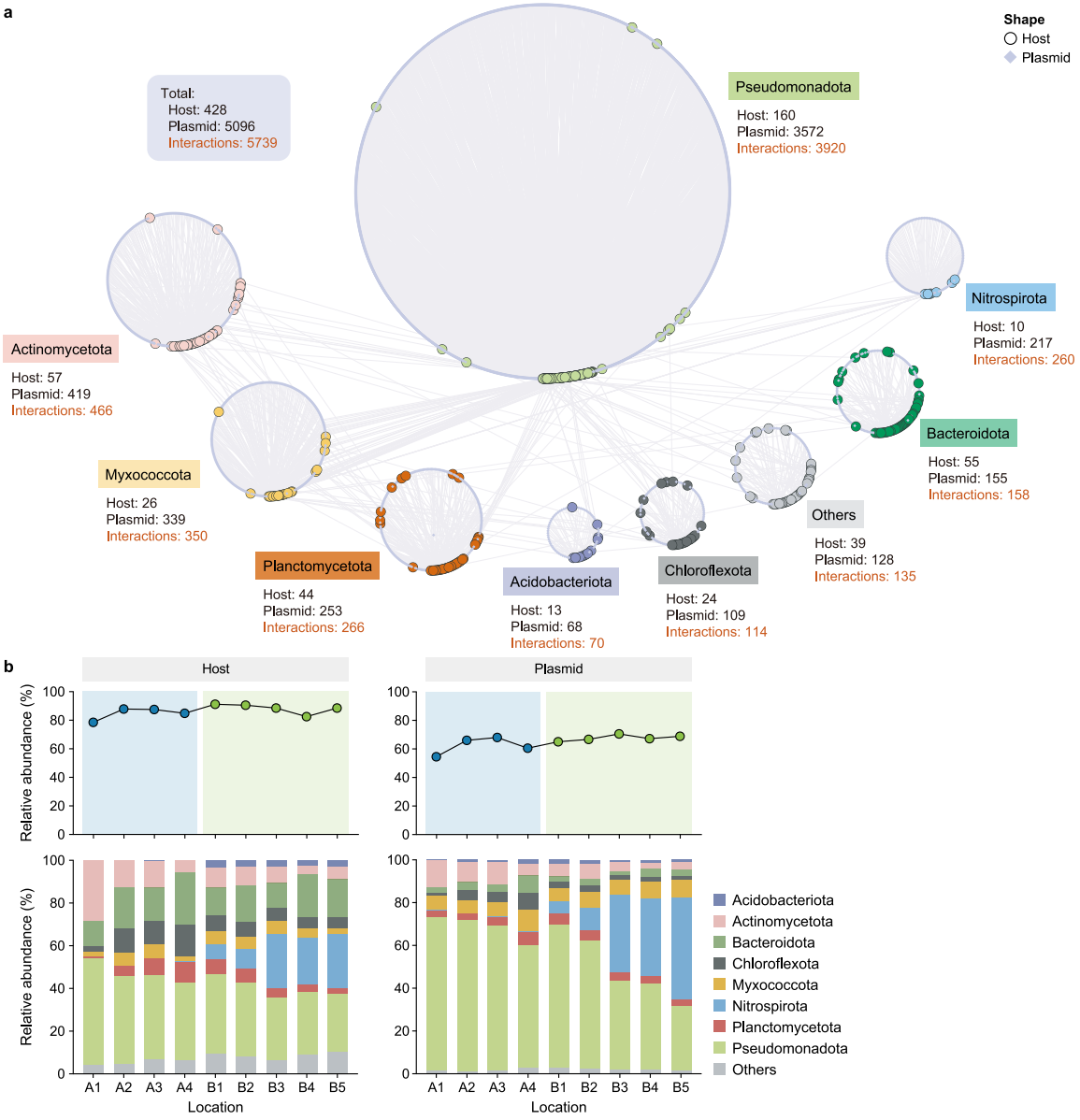

Fig. 5. Plasmid–host interaction network and community dynamics. **a,** Plasmid–host association network inferred using three methods: Hi-C, CRISPR spacer matching, and co-binning. Circular nodes represent prokaryotic hosts; lavender polygonal nodes denote plasmids. **b,** Phylum-level relative abundance of host organisms (left) and their associated plasmid (right) across sampling locations.

exceptionally high calculated plasmid carriage may be attributed to elevated plasmid fragmentation rates caused by high sample complexity, warranting further efforts to improve plasmid completeness.

3.6. Genetic mobility and adaptive potential enabled by plasmid-borne genes

Given the extensive plasmid–host interactions observed, we first tried to resolve plasmid transmission with biofilms. Metatranscriptomic data from each sampling site were analyzed to profile the expression patterns of genes involved in plasmid mobilization and conjugation. Consistent with plasmid–host associations, Pseudomonadota emerged as the dominant host phylum for plasmids harboring mobilization/conjugation potential. The expression of mobilization- and conjugation-related genes aligned with their coverage (Fig. 6a), with MOB genes exhibiting peak expression at location A1 (Supplementary Fig. S15), accounting for approximately 16.5% of the total expression of these genes (Supplementary Table S13). Notably, the expression of critical genes *traI*, *traM*, and *traY*, which are essential for initiating (e.g., nicking DNA and unwinding the DNA double helix) and facilitating plasmid conjugation, also peaked at A1. These results suggest that plasmid transfer was much more prevalent at this location.

To assess the functional relevance of the antibiotic resistance elements, the coverage and expression profiles of ARG-carrying contigs and plasmids were analyzed. Plasmid-borne ARGs demonstrated detectable coverage and expression across all locations (Fig. 6b). Notably, 9 of the 38 plasmid-associated ARGs colocalized with conjugation-related genes on 7 plasmids, and 29.6% (8/27) of the ARG-carrying plasmids exhibited multiple host associations revealed by our established plasmid–host linkages. These observations imply enhanced ARG horizontal dissemination via plasmid-mediated conjugation/mobilization. A total of 103 VFs were detected on 57 plasmids (Supplementary Table S14), including 8 VFs encoded by a single plasmid. Two ubiquitously expressed VFs, annotated as acyl-CoA dehydrogenase and RNA polymerase sigma factor, were linked to immune modulation and regulatory processes (Fig. 6c). Moreover, three VFs were colocalized with conjugation-related genes, suggesting potential horizontal transfer during plasmid transmission.

4. Discussion

Substantial numbers of phage and plasmid genomes were reconstructed from biofilms in a typical biofilm reactor using metaviromics and hybrid assembly, respectively, revealing their extensive diversity within this microbial community. Both the phage and plasmid communities exhibited strong stage-specific clustering across two concatenated RBC trains (Stages A and B) along the flow path. Moreover, the significant correlation between prokaryotic and phage/plasmid communities suggests that the distribution of these genetic elements is potentially driven by their prokaryotic hosts, consistent with theoretical expectations that these elements are highly host-dependent. In addition, shifts in the host community altered the abundance of these elements. For instance, shifts in microbial dominance along the flow path correlated with changes in phage coverage. The higher phage-to-host ratios in Stage A-dominated populations (derived from established host linkages) coincided with higher phage coverage, while the lower ratios in Stage B colonizers paralleled their decline, suggesting a dominance-driven phage–host balance (Supplementary Fig. S8).

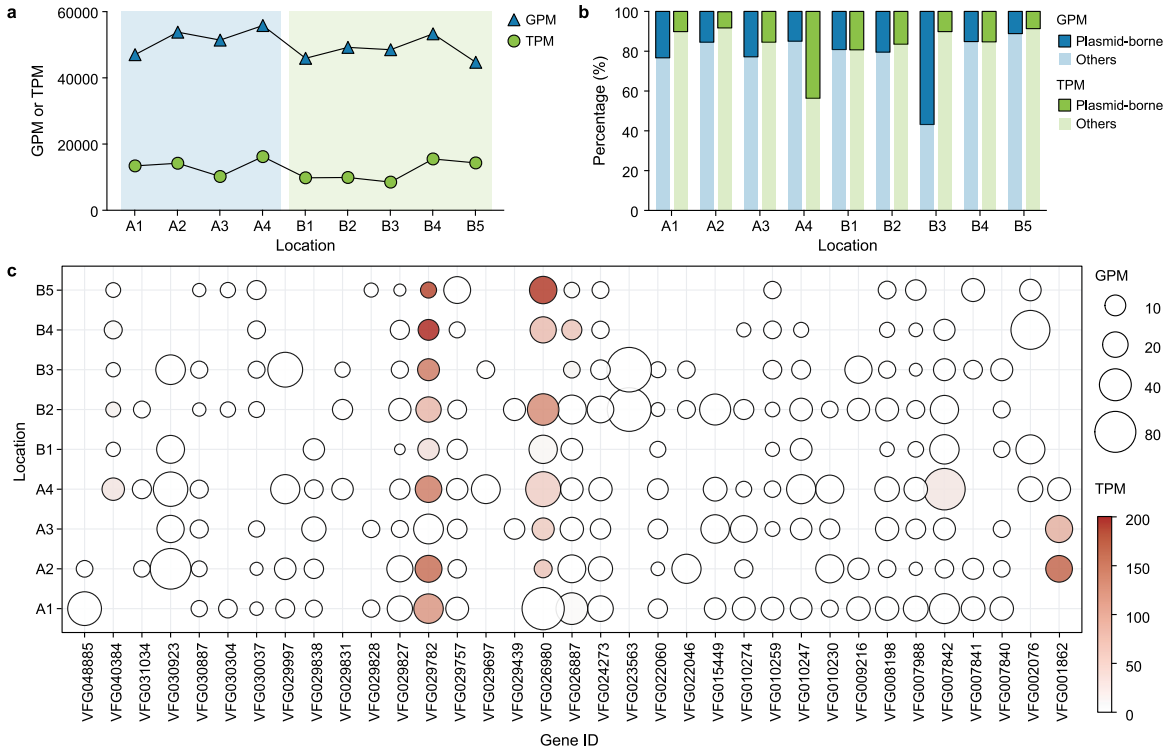

Fig. 6. Abundance and expression of functionally critical plasmid-associated genes. **a,** Coverage and expression of plasmid-encoded mobilization- and conjugation-associated genes. **b,** Comparative coverage and expression of antibiotic resistance genes (ARGs) across plasmids and chromosomal loci. **c,** Coverage and expression of virulence factors (VFs) harbored by plasmids. Gene IDs are referenced from the VF database. GPM, genes per million; TPM, transcripts per million.

In addition to the influence induced by the hosts of phages and plasmids, the biofilm architecture exerts multifaceted effects on the dynamics of these MGEs. For phages, biofilm thickness creates spatial heterogeneity in infection efficiency. Thin biofilms allow rapid phage penetration and host contact, but produced phages are likely to be released into the liquid phase rather than encountering new hosts for infection when biofilms are thinner and slow-growing [62], whereas thicker biofilms may impose diffusion barriers through dense extracellular polymeric substances (EPSs), which may physically trap phages or degrade them via EPS-associated nucleases. In the present study, Stage A showed higher overall phage coverage than Stage B. This could be related to the observed thinner biofilm layer at Stage B. For plasmids, biofilm thickness paradoxically influences plasmid transmission: while thickened biofilms theoretically enhance conjugation efficiency through increased cell density and opportunities for cell-to-cell contact [63,64], the dense EPS matrix in mature biofilms may simultaneously impair environmental DNA uptake and phage-mediated transduction. A recent investigation has shown that the spatial architecture of mature biofilms creates physical barriers that prevent donor cell infiltration into high-cell-density zones, thereby restricting direct donor-recipient contact and ultimately suppressing plasmid transfer efficiency. Conversely, early-stage biofilms (microcolonies) and those with incomplete surface colonization maintain accessible entry points for donor cells within regions that will later become densely packed microenvironments during maturation, thereby promoting successful plasmid transmission [14]. This finding was consistent with the observation in the present study that plasmid coverage declined at Stage A, where biofilms were thicker, and then increased at Stage B, where biofilms were thinner. To avoid disrupting the treatment facility's normal operation, we did not halt the RBC and therefore were unable to obtain detailed biofilm thickness data. Significantly, multiple factors may affect the colonization of phages and plasmids in biofilm systems, including diffusion, biofilm thickness and maturity, bacterial susceptibility, and the metabolic state of bacteria [65].

Additionally, the observed predominance of plasmids over phages in biofilm-derived metagenomic analyses likely stems from a combination of biological and methodological factors. Plasmids generally maintain higher intracellular copy numbers, and their persistence is enhanced within structured biofilm environments, where stable microbial communities promote long-term accumulation. In contrast, lysogenic phages typically persist as single-copy prophages within their host genomes. Lytic phages, while capable of rapid, high-titer intracellular replication, promptly lyse the host cell, releasing progeny into the environment. Furthermore, phage DNA, particularly from lytic phages, may be underrepresented. This is due to rapid degradation post-lysis, adsorption to EPSs, and challenges in releasing phage particles during standard DNA extraction protocols. Consequently, further investigation is warranted to elucidate the underlying mechanisms driving this disparity and to refine analytical approaches for capturing phage diversity in biofilms.

The present study systematically resolved phage– and plasmid–host interactions within biofilms using multiple approaches. Notably, as a complementary alternative to *in silico* methods, the proximity-ligation approach has tremendously expanded the elucidation of phage/plasmid–host interactions, and its performance in identifying host relationships has also been reported in recent studies [23,66,67]. In the present study, Hi-C data identified a greater number of plasmid–host pairs (5368 pairs; ~93% of the total) compared to phage–host pairs (429 pairs; ~81% of the total). Only a small fraction of these Hi-C linkages were corroborated by other methods (0.4% for plasmids and 9% for phages). This discrepancy likely stems from inherent biological differences in their integration, physical proximity to host DNA, and functional dynamics. Plasmids, which commonly exist as high-copy episomes, frequently interact with host chromosomes through physical colocalization at specific sites, such as replication hubs [68]. Their structural stability, intracellular abundance, and evolutionary reliance on host machinery likely facilitate efficient crosslinking and sustained chromosomal contact. In contrast, phage–host interactions are typically transient and culminate in host lysis. During late-stage infection, lytic phages lyse host cells, releasing fragmented host DNA and nonintegrated phage genomes. This material may escape Hi-C crosslinking, reducing detectable interactions. Moreover, many lytic phages deploy nucleases that actively degrade host chromatin during infection, repurposing nucleotides to amplify the burst size [69,70]. Such degradation fragments host DNA, further compromising Hi-C crosslinking efficiency. Critically, beyond inherent biological factors, the assembly quality of host and MGE genomes directly constrains the sensitivity of host identification. Thus, enhancing genome integrity is a priority for future methodological refinement. Nonetheless, Hi-C markedly improves the resolution of MGE–host identification within biofilms, offering precise clarity in mapping plasmid and phage associations within ecological networks.

In the studied biofilm system, phages likely function as dynamic regulators of bacterial populations through both lytic and lysogenic cycles. Consistent with this notion, restricted diffusion within biofilms has been proposed to enhance the coexistence of virulent phages and their hosts [62]—a scenario that aligns with the high abundance of virulent phages observed in the present study. Furthermore, we found that infected prokaryotes comprise a substantial portion of functional microbes through established phage–host associations, suggesting that phages may have important impacts on biogeochemical cycling by influencing microbial metabolic fluxes in the biofilm system (Supplementary Fig. S10). Additionally, phages might influence system carbon cycling by expressing AMGs, including genes involved in the breakdown of complex carbohydrates, to boost energy production. These include glycoside hydrolases and polysaccharide lyases, which facilitate the utilization of polysaccharide chains, such as starch, xylan, and rhamnose. Phage-encoded glycoside hydrolases have also been found in various habitats, including deep-sea sediments, peatland soils, and bovine rumen [4,52,71]. Moreover, phages might enhance carbon fixation by expressing the AMGs involved in the reductive citrate cycle and the reductive pentose phosphate cycle [72]. Although phages have been implicated in the transfer of ARGs in several studies [73,74], no ARGs were identified in any reconstructed phages in the present study. Similarly, no VFs were detected in the phages.

In contrast, plasmids may function as mobile genetic “toolkits” that could accelerate biofilm adaptation and gene transfer. Their role as vectors for ARGs, VFs, and metabolic genes enables the rapid dissemination of adaptive traits across densely structured biofilm communities [75]. In the present study, both ARGs and VFs were identified and expressed in plasmids identified from the biofilm. Consistent with other research, plasmid–host linkage analysis confirmed a broad host range, suggesting the potential of plasmids to disseminate ARGs within the biofilm. The prevalence of hosts harboring these ARG-carrying plasmids identified microbial taxa that may require further attention to mitigate potential risks.

The synergistic and antagonistic interplay among MGEs is likely central to biofilm resilience and ecological adaptation. Notably, in the present study, more than one-third (184/525) of the prokaryotic hosts were linked to plasmids and phages, suggesting the tight

interconnection of these MGEs within the biofilm. Such connectivity suggests potential coevolutionary dynamics: phages may facilitate the horizontal transfer of plasmid-encoded genes, such as ARGs or AFs, through transduction or by releasing plasmid DNA during lysis, thereby potentially enhancing host fitness under specific conditions. Conversely, plasmids may encode antiphage defense systems (e.g., restriction-modification systems) to resist infection [76] and prevent phage invasion, while phages may in turn selectively lyse hosts carrying plasmids with costly traits, indirectly modulating plasmid prevalence [77]. Understanding these complex interactions is essential to deciphering biofilm-associated challenges—from the dissemination of antibiotic resistance to the optimization of bioremediation strategies.

5. Conclusion

In conclusion, integrating multi-omics and Hi-C methods provides a holistic framework for unraveling the assembly dynamics and interaction networks of phages and plasmids with their prokaryotic hosts. The present study successfully reconstructed a diverse array of vOTUs and plasmids, revealing their strong spatial and functional associations with microbial communities across the biofilm reactor. Some encoded exogenous genes likely underpin the roles of phages and plasmids in modulating biofilm structure, functional potential, and environmental adaptability. Although advanced computational methods have greatly improved the identification of phage–host and plasmid–host connections, a substantial fraction of these elements remains unlinked to specific hosts, underscoring the ongoing challenges of deciphering complex microbial interactions. Future research integrating experimental validation and novel bioinformatic tools is needed to fully elucidate the ecological and evolutionary significance of phages and plasmids in biofilm ecosystems.

CRediT authorship contribution statement

Dou Wang: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Visualization. **Xiaoqing Xu:** Methodology, Investigation. **Lei Liu:** Software, Methodology. **Chunxiao Wang:** Methodology, Investigation. **Yu Deng:** Investigation, Methodology. **Martin F. Polz:** Writing – review & editing. **Tong Zhang:** Writing – review & editing, Supervision, Resources, Conceptualization, Methodology.

Data availability

All sequencing data generated in this study and recovered MAGs are deposited in the NCBI database under the BioProject ID PRJNA974210.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.

Acknowledgements

Dr. Dou Wang and Dr. Xiaoqing Xu would like to thank The University of Hong Kong for the postdoctoral fellowship. Technical assistance from Ms. Vicky Fung is greatly appreciated. This work was financially supported by the General Research Fund (17212124) of Hong Kong.

Appendix. ASupplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ese.2026.100683.

References

[1] S.M. Arbas, S. Narayanasamy, M. Herold, L.A. Lebrun, M.R. Hoopmann, S. Li, T.J. Lam, B.J. Kunath, N.D. Hicks, C.M. Liu, Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics, Nat. Microbiol. 6 (2020) 1–13.
[2] A.P. Camargo, S. Roux, F. Schulz, M. Babinski, Y. Xu, B. Hu, P.S.G. Chain, S. Nayfach, N.C. Kyrpides, Identification of mobile genetic elements with geNomad, Nat. Biotechnol. 42 (8) (2024) 1303–1312.
[3] S. Redondo-Salvo, R. Fernández-López, R. Ruiz, L. Vielva, M. de Toro, E.P.C. Rocha, M.P. Garcillán-Barcia, F. de la Cruz, Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids, Nat. Commun. 11 (1) (2020) 3602.
[4] Z. Li, D. Pan, G. Wei, W. Pi, C. Zhang, J.-H. Wang, Y. Peng, L. Zhang, Y. Wang, C.R.J. Hubert, X. Dong, Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity, ISME J. 15 (8) (2021) 2366–2378.
[5] C. Howard-Varona, M.M. Lindback, G.E. Bastien, N. Solonenko, A.A. Zayed, H. Jang, B. Andreopoulos, H.M. Brewer, T. Glavina del Rio, J.N. Adkins, S. Paul, M.B. Sullivan, M.B. Duhaime, Phage-specific metabolic reprogramming of virocells, ISME J. 14 (4) (2020) 881–895.
[6] P. Mara, D. Vik, M.G. Pachiadaki, E.A. Suter, B. Poulos, G.T. Taylor, M.B. Sullivan, V.P. Edgcomb, Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline, ISME J. 14 (12) (2020) 3079–3092.
[7] F. Benz, A.R. Hall, Host-specific plasmid evolution explains the variable spread of clinical antibiotic-resistance plasmids, Proc. Natl. Acad. Sci. 120 (15) (2023) e2212147120.
[8] B. Pakbin, W.M. Brück, J.W. Rossen, Virulence factors of enteric pathogenic Escherichia coli: a review, Int. J. Mol. Sci. 22 (18) (2021) 9922.
[9] D. Dai, C. Brown, H. Bürgmann, D.G.J. Larsson, I. Nambi, T. Zhang, C.-F. Flach, A. Pruden, P.J. Vikesland, Long-read metagenomic sequencing reveals shifts in associations of antibiotic resistance genes with mobile genetic elements from sewage to activated sludge, Microbiome 10 (1) (2022) 20.
[10] Y. Chen, Y. Wang, D. Paez-Espino, M.F. Polz, T. Zhang, Prokaryotic viruses impact functional microorganisms in nutrient removal and carbon cycle in wastewater treatment plants, Nat. Commun. 12 (1) (2021) 5398.
[11] Y. Che, Y. Xia, L. Liu, A.-D. Li, Y. Yang, T. Zhang, Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing, Microbiome 7 (2019) 1–13.
[12] E. Spasov, J.M. Tsuji, L.A. Hug, A.C. Doxey, L.A. Sauder, W.J. Parker, J.D. Neufeld, High functional diversity among Nitrospira populations that dominate rotating biological contactor microbial communities in a municipal wastewater treatment plant, ISME J. 14 (7) (2020) 1857–1872.
[13] D. Freeman, Y.B. Fernández, A. Wilson, B.A. McKew, C. Whitby, D.R. Clark, B. Jefferson, F. Coulon, F. Hassard, Nitrogen oxidation consortia dynamics influence the performance of full-scale rotating biological contactors, Environ. Int. 135 (2020) 105354.
[14] S. Djermoun, D.K. Rode, E. Jiménez-Siebert, N. Netter, C. Lesterlin, K. Drescher, S. Bigot, Biofilm architecture determines the dissemination of conjugative plasmids, Proc. Natl. Acad. Sci. 122 (17) (2025) e2417452122.
[15] E.V. Koonin, M. Krupovic, Evolution of adaptive immunity from transposable elements combined with innate immune systems, Nat. Rev. Genet. 16 (3) (2015) 184–192.
[16] K.S. Makarova, Y.I. Wolf, O.S. Alkhnbashi, F. Costa, S.A. Shah, S.J. Saunders, R. Barrangou, S.J. Brouns, E. Charpentier, D.H. Haft, An updated evolutionary classification of CRISPR–Cas systems, Nat. Rev. Microbiol. 13 (11) (2015) 722–736.
[17] R.A. Edwards, K. McNair, K. Faust, J. Raes, B.E. Dutilh, Computational approaches to predict bacteriophage–host relationships, FEMS Microbiol. Rev. 40 (2) (2016) 258–272.
[18] J. Shang, Y. Sun, CHERRY: a computational metHod for accuratE pRediction of virus–pRokarYotic interactions using a graph encoder–decoder model, Briefings Bioinf. 23 (5) (2022).
[19] Y. Ji, J. Shang, X. Tang, Y. Sun, HOTSPOT: hierarchical host prediction for assembled plasmid contigs with transformer, Bioinformatics 39 (5) (2023).
[20] Y. Zhang, B. Xue, Y. Mao, X. Chen, W. Yan, Y. Wang, Y. Wang, L. Liu, J. Yu, X. Zhang, S. Chao, E. Topp, W. Zheng, T. Zhang, High-throughput single-cell sequencing of activated sludge microbiome, Environ. Sci. Ecotechnol. 23 (2025) 100493.
[21] K. Arikawa, K. Ide, M. Kogawa, T. Saeki, T. Yoda, T. Endoh, A. Matsuhashi, H. Takeyama, M. Hosokawa, Recovery of strain-resolved genomes from human microbiome through an integration framework of single-cell genomics and metagenomics, Microbiome 9 (1) (2021) 202.
[22] G. Uritskiy, M. Press, C. Sun, G.D. Huerta, A.A. Zayed, A. Wiser, J. Grove,

B. Auch, S.M. Eacker, S. Sullivan, D.M. Bickhart, T.P.L. Smith, M.B. Sullivan, I. Liachko, Accurate viral genome reconstruction and host assignment with proximity-ligation sequencing, bioRxiv (2021). https://doi.org/10.1101/2021.06.14.448389.
[23] T. Stalder, M.O. Press, S. Sullivan, I. Liachko, E.M. Top, Linking the resistome and plasmidome to the microbiome, ISME J. 13 (10) (2019) 2437–2446.
[24] A.G. Kent, A.C. Vill, Q. Shi, M.J. Satlin, I.L. Brito, Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C, Nat. Commun. 11 (1) (2020) 4379.
[25] R. Wu, M.R. Davison, W.C. Nelson, M.L. Smith, M.S. Lipton, J.K. Jansson, R.S. McClure, J.E. McDermott, K.S. Hofmockel, Hi-C metagenome sequencing reveals soil phage–host interactions, Nat. Commun. 14 (1) (2023) 7666.
[26] N.S. Varona, P.J. Hesketh-Best, F.H. Coutinho, A.K. Stiffler, B.A. Wallace, S.L. Garcia, Y. Scholten, A.F. Haas, M. Little, M. Vermeij, A. Luque, C. Silveira, Host-specific viral predation network on coral reefs, ISME J. 18 (1) (2024).
[27] D. Wang, Y. Wang, L. Liu, Y. Chen, C. Wang, X. Xu, Y. Yang, Y. Wang, T. Zhang, Niche differentiation and symbiotic association among ammonia/nitrite oxidizers in a full-scale rotating biological contactor, Water Res. 225 (2022) 119137.
[28] S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) i884–i890.
[29] L. Liu, Y. Wang, Y. Yang, D. Wang, S.H. Cheng, C. Zheng, T. Zhang, Charting the complexity of the activated sludge microbiome through a hybrid sequencing strategy, Microbiome 9 (1) (2021) 205.
[30] M.R. Olm, C.T. Brown, B. Brooks, J.F. Banfield, dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication, ISME J. 11 (12) (2017) 2864–2868.
[31] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, Genome Res. 25 (7) (2015) 1043–1055.
[32] S. Nurk, D. Meleshko, A. Korobeynikov, P.A. Pevzner, metaSPAdes: a new versatile metagenomic assembler, Genome Res. 27 (5) (2017) 824–834.
[33] D. Bertrand, J. Shaw, M. Kalathiyappan, A.H.Q. Ng, M.S. Kumar, C. Li, M. Dvornicic, J.P. Soldo, J.Y. Koh, C. Tong, O.T. Ng, T. Barkham, B. Young, K. Marimuthu, K.R. Chng, M. Sikic, N. Nagarajan, Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes, Nat. Biotechnol. 37 (8) (2019) 937–944.
[34] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C.A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS One 9 (11) (2014) e112963.
[35] P.S. Krawczyk, L. Lipinski, A. Dziembowski, PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures, Nucleic Acids Res. 46 (6) (2018) e35-e35.
[36] D. Hyatt, G.-L. Chen, P.F. LoCascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification, BMC Bioinf. 11 (1) (2010) 119.
[37] X. Yin, X. Zheng, L. Li, A.-N. Zhang, X.-T. Jiang, T. Zhang, ARGs-OAP v3.0: Antibiotic-Resistance gene database curation and analysis pipeline optimization, Engineering 27 (2023) 234–241.
[38] S. Zhou, B. Liu, D. Zheng, L. Chen, J. Yang, Vfdb 2025: an integrated resource for exploring anti-virulence compounds, Nucleic Acids Res. 53 (D1) (2024) D871–D877.
[39] J. Guo, B. Bolduc, A.A. Zayed, A. Varsani, G. Dominguez-Huerta, T.O. Delmont, A.A. Pratama, M.C. Gazitúa, D. Vik, M.B. Sullivan, S. Roux, VirSorter2: a multiclassifier, expert-guided approach to detect diverse DNA and RNA viruses, Microbiome 9 (1) (2021) 37.
[40] D. Paez-Espino, G.A. Pavlopoulos, N.N. Ivanova, N.C. Kyrpides, Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data, Nat. Protoc. 12 (8) (2017) 1673–1682.
[41] S. Roux, E.M. Adriaenssens, B.E. Dutilh, E.V. Koonin, A.M. Kropinski, M. Krupovic, J.H. Kuhn, R. Lavigne, J.R. Brister, A. Varsani, Minimum information about an uncultivated virus genome (MIUViG), Nat. Biotechnol. 37 (1) (2019) 29–37.
[42] S. Nayfach, A.P. Camargo, F. Schulz, E. Eloe-Fadrosh, S. Roux, N.C. Kyrpides, CheckV assesses the quality and completeness of metagenome-assembled viral genomes, Nat. Biotechnol. 39 (5) (2021) 578–585.
[43] H.B. Jang, B. Bolduc, O. Zablocki, J.H. Kuhn, S. Roux, E.M. Adriaenssens, J.R. Brister, A.M. Kropinski, M. Krupovic, R. Lavigne, Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks, Nat. Biotechnol. 37 (6) (2019) 632–639.
[44] J.-Z. Jiang, W.-G. Yuan, J. Shang, Y.-H. Shi, L.-L. Yang, M. Liu, P. Zhu, T. Jin, Y. Sun, L.-H. Yuan, Virus classification for viral genomic fragments using PhaGCN2, Briefings Bioinf. 24 (1) (2022) bbac505.
[45] M. Shaffer, M.A. Borton, B.B. McGivern, A.A. Zayed, Sabina L. La Rosa, L.M. Solden, P. Liu, A.B. Narrowe, J. Rodríguez-Ramos, B. Bolduc, M.C. Gazitúa, R.A. Daly, G.J. Smith, D.R. Vik, P.B. Pope, M.B. Sullivan, S. Roux, Kelly C. Wrighton, DRAM for distilling microbial metabolism to automate the curation of microbiome function, Nucleic Acids Res. 48 (16) (2020) 8883–8900.
[46] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (6) (2015) 845–858.
[47] A.J. Hockenberry, C.O. Wilke, BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains, PeerJ 9 (2021) e11396.
[48] J. Shang, X. Tang, Y. Sun, PhaTYP: predicting the lifestyle for bacteriophages using BERT, Briefings Bioinf. 24 (1) (2023) bbac487.
[49] C. Bland, T.L. Ramsey, F. Sabree, M. Lowe, K. Brown, N.C. Kyrpides, P. Hugenholtz, CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats, BMC Bioinf. 8 (1) (2007) 209.
[50] S. Nayfach, S. Roux, R. Seshadri, D. Udwary, N. Varghese, F. Schulz, D. Wu, D. Paez-Espino, I.M. Chen, M. Huntemann, K. Palaniappan, J. Ladau, S. Mukherjee, T.B.K. Reddy, T. Nielsen, E. Kirton, J.P. Faria, J.N. Edirisinghe, C.S. Henry, S.P. Jungbluth, D. Chivian, P. Dehal, E.M. Wood-Charlson, A.P. Arkin, S.G. Tringe, A. Visel, I.M.D. Consortium, T. Woyke, N.J. Mouncey, N.N. Ivanova, N.C. Kyrpides, E.A. Eloe-Fadrosh, A genomic catalog of Earth's microbiomes, Nat. Biotechnol. 39 (4) (2021) 499–509.
[51] S. Roux, J.R. Brum, B.E. Dutilh, S. Sunagawa, M.B. Duhaime, A. Loy, B.T. Poulos, N. Solonenko, E. Lara, J. Poulain, S. Pesant, S. Kandels-Lewis, C. Dimier, M. Picheral, S. Searson, C. Cruaud, A. Alberti, C.M. Duarte, J.M. Gasol, D. Vaqué, P. Bork, S.G. Acinas, P. Wincker, M.B. Sullivan, Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses, Nature 537 (7622) (2016) 689–693.
[52] J.B. Emerson, S. Roux, J.R. Brum, B. Bolduc, B.J. Woodcroft, H.B. Jang, C.M. Singleton, L.M. Solden, A.E. Naas, J.A. Boyd, Host-linked soil viral ecology along a permafrost thaw gradient, Nat. Microbiol. 3 (8) (2018) 870–880.
[53] P.P. Chan, B.Y. Lin, A.J. Mak, T.M. Lowe, tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes, Nucleic Acids Res. 49 (16) (2021) 9077–9096.
[54] D. Paez-Espino, E.A. Eloe-Fadrosh, G.A. Pavlopoulos, A.D. Thomas, M. Huntemann, N. Mikhailova, E. Rubin, N.N. Ivanova, N.C. Kyrpides, Uncovering Earth's virome, Nature 536 (7617) (2016) 425–430.
[55] E. Kopylova, L. Noé, H.J.B. Touzet, SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data, Bioinformatics 28 (24) (2012) 3211–3217.
[56] R. Niederdorfer, D. Hausherr, A. Palomo, J. Wei, P. Magyar, B.F. Smets, A. Joss, H. Bürgmann, Temperature modulates stress response in mainstream anammox reactors, Commun. Biol. 4 (1) (2021) 23.
[57] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinf. 12 (1) (2011) 323.
[58] S.T.N. Aroney, R.J.P. Newell, J. Nissen, A.P. Camargo, G.W. Tyson, B.J. Woodcroft, Coverm: Read Coverage Calculator for Metagenomics, 2024.
[59] R. R Core Team, R: a Language and Environment for Statistical Computing, 2013.
[60] A.A. Pratama, B. Bolduc, A.A. Zayed, Z.-P. Zhong, J. Guo, D.R. Vik, M.C. Gazitúa, J.M. Wainaina, S. Roux, M.B. Sullivan, Expanding standards in viromics: in silico evaluation of dsDNA viral genome identification, classification, and auxiliary metabolic gene curation, PeerJ 9 (2021) e11447.
[61] A.M. ter Horst, C. Santos-Medellín, J.W. Sorensen, L.A. Zinke, R.M. Wilson, E.R. Johnston, G.G. Trubl, J. Pett-Ridge, S.J. Blazewicz, P.J. Hanson, Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations, Microbiome 9 (1) (2021) 1–18.
[62] M. Simmons, K. Drescher, C.D. Nadell, V. Bucci, Phage mobility is a core determinant of phage-bacteria coexistence in biofilms, ISME J. 12 (2) (2018) 531–543.
[63] B. Normander, B.B. Christensen, S. Molin, N. Kroer, Effect of bacterial distribution and activity on conjugal gene transfer on the phylloplane of the bush bean (Phaseolus vulgaris), Appl. Environ. Microbiol. 64 (5) (1998) 1902–1909.
[64] S. Molin, T. Tolker-Nielsen, Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure, Curr. Opin. Biotechnol. 14 (3) (2003) 255–261.
[65] C. Michaelis, E. Grohmann, Horizontal gene transfer of antibiotic resistance genes in biofilms, Antibiotics 12 (2) (2023) 328.
[66] D.M. Bickhart, M. Kolmogorov, E. Tseng, D.M. Portik, A. Korobeynikov, I. Tolstoganov, G. Uritskiy, I. Liachko, S.T. Sullivan, S.B. Shin, A. Zorea, V.P. Andreu, K. Panke-Buisse, M.H. Medema, I. Mizrahi, P.A. Pevzner, T.P.L. Smith, Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities, Nat. Biotechnol. 40 (5) (2022) 711–719.
[67] M. Marbouty, A. Thierry, G.A. Millot, R. Koszul, MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut, eLife 10 (2021) e60608.
[68] R.T. Wheeler, L. Shapiro, Bacterial chromosome segregation: is there a mitotic apparatus? Cell 88 (5) (1997) 577–579.
[69] A.C. McKitterick, S.G. Hays, F.-T. Johura, M. Alam, K.D. Seed, Viral satellites exploit phage proteins to escape degradation of the bacterial host chromosome, Cell Host Microbe 26 (4) (2019) 504–514, e504.
[70] J.Y. Yang, W. Fang, F. Miranda-Sanchez, J.M. Brown, K.M. Kauffman, C.M. Acevero, D.P. Bartel, M.F. Polz, L. Kelly, Degradation of host translational machinery drives tRNA acquisition in viruses, Cell Syst. 12 (8) (2021) 771–779, e775.
[71] C.L. Anderson, M.B. Sullivan, S.C. Fernando, Dietary energy drives the dynamic response of bovine rumen viral communities, Microbiome 5 (1) (2017) 155.
[72] F.H. Coutinho, P.J. Cabello-Yeves, R. Gonzalez-Serrano, R. Rosselli, M. López-Pérez, T.I. Zemskaya, A.S. Zakharenko, V.G. Ivanov, F. Rodriguez-Valera, New viral biogeochemical roles revealed through metagenomic analysis of Lake Baikal, Microbiome 8 (1) (2020) 163.

[73] R. Cook, S. Hooton, U. Trivedi, L. King, C.E.R. Dodd, J.L. Hobman, D.J. Stekel, M.A. Jones, A.D. Millard, Hybrid assembly of an agricultural slurry virome reveals a diverse and stable community with the potential to alter the metabolism and virulence of veterinary pathogens, Microbiome 9 (1) (2021) 65.
[74] W. Calero-Cáceres, M. Ye, J.L. Balcázar, Bacteriophages as environmental reservoirs of antibiotic resistance, Trends Microbiol. 27 (7) (2019) 570–577.
[75] J.S. Madsen, M. Burmølle, L.H. Hansen, S.J. Sørensen, The interconnection between biofilm formation and horizontal gene transfer, FEMS Immunol. Med. Microbiol. 65 (2) (2012) 183–195.
[76] D.M. Picton, Y.A. Luyten, R.D. Morgan, A. Nelson, D.L. Smith, D.T. Dryden, J.C. Hinton, T.R. Blower, The phage defence island of a multidrug resistant plasmid uses both BREX and type IV restriction for complementary protection from viruses, Nucleic Acids Res. 49 (19) (2021) 11257–11273.
[77] E. Harrison, A.J. Wood, C. Dytham, J.W. Pitchford, J. Truman, A. Spiers, S. Paterson, M.A. Brockhurst, Bacteriophages limit the existence conditions for conjugative plasmids, mBio 6 (3) (2015).