Liu *et al. Microbiome* (2025) 13:14
https://doi.org/10.1186/s40168-024-02019-0

Microbiome

RESEARCH

Open Access

Rumen DNA virome and its relationship with feed efficiency in dairy cows

Xiaohan Liu[1], Yifan Tang[1], Hongyi Chen[1], Jian-Xin Liu[1] and Hui-Zeng Sun[1*]

Abstract

Background The rumen harbors a diverse virome that interacts with other microorganisms, playing pivotal roles in modulating metabolic processes within the rumen environment. However, the characterization of rumen viruses remains incomplete, and their association with production traits, such as feed efficiency (FE), has not been documented. In this study, rumen fluid from 30 Chinese Holstein dairy cows was analyzed using next-generation sequencing (NGS) and High-Fidelity (HiFi) sequencing to elucidate the rumen DNA virome profile and uncover potential viral mechanisms influencing FE.

Results Integrated NGS and HiFi sequencing enhanced the length, completeness, and resolution of viral operational taxonomic units (vOTUs) compared to NGS. A total of 6,922 vOTUs were identified, including 4,716 lytic and 1,961 temperate vOTUs. At the family level, lytic viruses were predominantly from Siphoviridae (30.35%) and Schitoviridae (23.93%), while temperate viruses were primarily Siphoviridae (67.21%). The study annotated 2,382 auxiliary metabolic genes (AMGs), comprising 1,752 lytic virus-associated AMGs across 51 functional categories and 589 temperate virus-associated AMGs across 29 categories. Additionally, 2,232 vOTU-host metagenome-assembled genome (hMAG) linkages were predicted, with Firmicutes_A (33.60%) and Bacteroidota (33.24%) being the most prevalent host phyla. Significant differences in viral populations were observed between high and low FE groups across multiple taxonomic levels ($P < 0.05$). Two pathways were proposed to explain how rumen viruses might modulate FE: (1) Lytic viruses could lyse beneficial host bacteria linked to favorable cattle phenotypes, such as vOTU1836 targeting Ruminococcaceae, resulting in diminished organic acid production and consequently lower FE; (2) AMG-mediated host metabolism modulation, exemplified by *GT2* carried by vOTU0897, which may enhance Lachnospiraceae fermentation capacity, increasing organic acid production and thereby improving FE.

Conclusions This study constructed a comprehensive rumen DNA virome profile for Holstein dairy cows, elucidating the structural and functional complexity of rumen viruses, the roles of AMGs, and vOTU-hMAG linkages. The integration of these data offers novel insights into the mechanisms by which rumen viruses may regulate nutrient utilization, potentially influencing FE in dairy cows.

Keywords HiFi sequencing, Lytic, Temperate, Auxiliary metabolic genes, Feed efficiency

Introduction

Enhancing feed efficiency (FE) in dairy cows is a critical objective to promote the sustainable development of the dairy industry [1]. The rumen hosts a vast and diverse microbial ecosystem encompassing multiple kingdoms, which play a pivotal role in the degradation of plant biomass, thereby supplying a substantial portion of nutrient precursors essential for dairy cow productivity and,

*Correspondence:
Hui-Zeng Sun
huizeng@zju.edu.cn
[1] Institute of Dairy Science, MoE Key Laboratory of Molecular Animal Nutrition, College of Animal Sciences, Zhejiang University, Hangzhou, China

consequently, influencing FE. The majority of existing research has concentrated on ruminal bacteria and archaea [2–5]. For instance, positive synergies between *Selenomonas* species and members of the *Succinivibrionaceae* family have been correlated with higher FE in dairy cows [2]. Additionally, some investigations have explored the roles of ruminal fungi and protozoa [6, 7]. However, despite the critical contribution of ruminal microorganisms to dairy cow phenotypic traits, the study of rumen viruses remains relatively underexplored. Viruses are the most abundant acellular entities globally [8], with the density of free viral particles in the rumen ranging from 10^7 to 10^10 particles per milliliter [9–12], highlighting a substantial yet inadequately studied population in the rumen. Additionally, viruses parasitize all other microbial kingdoms, modulating the rumen microbiome and thereby impacting animal productivity [13, 14], which indicates that viruses play a vital role in the rumen ecosystem.

Viruses in the rumen can be broadly categorized based on their life cycle into two main types: lytic and temperate. Lytic viruses infect host cells, rapidly replicate, and induce cell lysis, leading to the release of progeny virions that can infect additional hosts [15]. In contrast, temperate viruses integrate their genetic material into the host genome, establishing the provirus stage, which enables a prolonged symbiotic relationship with the host [16]. Later, the prophage can be induced into the lytic cycle either spontaneously or by physical or chemical factors [17]. These distinct viral infection strategies impart unique functional roles within the host cells. Viruses can modulate host metabolism through various mechanisms, mainly including the lysis of host cells [18], and the presence of auxiliary metabolic genes (AMGs) that can selectively influence host metabolic processes [19, 20]. The lysis induced by viruses not only leads to the immediate demise of the infected cells but also causes a reduction in the host community numbers over a certain period [21]. AMGs encompass a broad spectrum of genes that regulate critical host metabolic and homeostatic functions [22], participating in numerous metabolic pathways, such as photosynthesis, phosphate metabolism, central carbon metabolism, nutrient cycling, and nucleotide biosynthesis [19, 23, 24]. This combination of lytic effects and metabolic manipulation allows viruses to play a significant role in shaping the dynamics and functionality of the rumen microbiome.

Research on rumen viruses in ruminants has evolved through three key phases: morphological examination, molecular biology investigation, and contemporary omics-based studies. As early as the 1860s, the presence of viruses in the rumen was documented [25, 26]. Morphological studies elucidated the structures of several viral families, including *Siphoviridae*, *Podoviridae*, *Myoviridae*, and various non-tailed phages [27]. Subsequent molecular biology research primarily focused on the genomic characterization of bacteriophages, including genome length analysis and restriction endonuclease patterns. Specific bacteriophages targeting rumen-specific bacterial species, such as *Streptococcus bovis* and *Selenomonas ruminantium*, were identified and isolated [28, 29]. The advent of omics-based research has facilitated a more comprehensive identification and functional characterization of these viruses [12, 30–33]. For example, metagenomic sequencing enabled the identification of 14 core rumen viral groups and revealed that AMGs participate in diverse metabolic pathways [31]. Integrative viromics and proteomics approaches have allowed the assembly of extensive viral contigs, providing insights into the interactions between rumen viruses and dominant carbohydrate-degrading microorganisms, with implications for gastrointestinal carbon cycling [32]. Notably, the creation of the Global Rumen Virome Database (RVD) [33] and the Unified Ruminant Phage Catalogue (URPC) [34] have offered significant insights into viral diversity, host interactions, and potential modulation of rumen functions. However, these studies often rely on next-generation sequencing (NGS) technologies, which, despite their utility, yield short read lengths that require assembly and are susceptible to errors, particularly in regions with high sequence repetition or structural variations [35].

The emergence of third-generation sequencing technologies, such as nanopore sequencing from Oxford Nanopore Technologies and Single Molecule Real Time (SMRT) sequencing from Pacific Biosciences (PacBio), has overcome many of the limitations associated with NGS by enabling the generation of extended read lengths [36, 37]. Initially, these technologies produced long reads with lower accuracy, necessitating error correction using high-precision Illumina short reads [38, 39]. However, advances such as PacBio's HiFi sequencing now allow for the production of highly accurate long reads, with substantial implications for genome assembly and metagenomic studies [40]. Although HiFi sequencing has significantly advanced the quality of human genome and gut microbial metagenome assembled genomes (MAGs) [40, 41], its application to the study of bovine rumen viruses remains largely unexplored.

Given this context, we hypothesize that viruses affect the FE of dairy cows by influencing the structure and functional composition of their hosts, achieved through two pathways: 1) The direct lysis of host cells associated with production traits, thereby reducing

the abundance of these hosts and impacting FE; and 2) The modulation of host metabolism by AMGs, which may enhance or diminish the production of metabolites linked to FE, affecting it accordingly. To test this hypothesis, we assessed FE in 53 mid-lactation Holstein dairy cows and selected 15 high-efficiency (HE) and 15 low-efficiency (LE) cows. We then performed NGS ($n=30$) and HiFi ($n=1$) sequencing to construct a comprehensive rumen DNA virome profile, including viral taxonomy, composition, diversity, virus-host linkages, and information on AMGs carried by the viruses. By integrating these data and identifying differential viruses between the HE and LE groups as "biomarkers," we aim to preliminarily explore the potential roles of these viruses in modulating FE.

Results

Animal phenotypes and metagenomic data

Energy-corrected milk (ECM, $P=0.0003$), milk yield ($P=0.0002$), and ECM/DMI (dry matter intake, 1.593 ± 0.017 vs. 1.374 ± 0.019, $P<0.0001$) were significantly elevated in the HE group (Table 1; additional details in Table S1). Conversely, there were no statistically significant differences in DMI ($P=0.5216$), milk protein content ($P=0.1280$), and milk fat content ($P=0.9918$) between the HE and LE groups (Table 1; further details in Table S1).

NGS generated an average of 40.51 ± 2.04 Gb (mean ± SEM) of data per sample. A total of 8,102,834,056 raw reads were produced from the NGS analysis of 30 rumen fluid samples, with an average of $270{,}094{,}468 \pm 13{,}573{,}491$ reads per sample (Table S2). After implementing quality control measures and removing host gene sequences, 7,974,091,852 clean reads were retained, averaging $265{,}803{,}062 \pm 13{,}460{,}468$ reads per sample (Table S3). The HiFi sequencing for one rumen fluid sample produced 15.60 Gb of data with an average read length of 6.03 Kb (more details in Table S4).

Table 1 Milk production and feed efficiency in cows selected for HE and LE

Items	Groups		SEM	*P*-value
	LE	HE		
DMI (kg/d)	25.030	26.701	0.951	0.5216
Milk Yield (kg)	30.560	36.416	1.393	0.0002
Milk fat content (%)	4.143	4.145	0.193	0.9918
Milk protein content (%)	3.377	3.279	0.062	0.1280
ECM	33.348	39.600	1.534	0.0003
ECM/DMI	1.374	1.593	0.025	< 0.0001

LE Low feed efficiency, *HE* High feed efficiency, *DMI* Dry matter intake, *ECM* Energy corrected milk. ECM = (0.3246 × kg of milk) + (13.86 × kg of milk fat) + (7.04 × kg of milk protein)

Overview of rumen DNA virome

To evaluate the advantages of the integration of NGS and HiFi sequencing (NGS + HiFi) over NGS in rumen virome analysis, we examined two primary aspects: the number of viral operational taxonomic units (vOTUs) and their average length. As the contig length threshold for filtration increased (2k, 5k, 8k, 10k, and15 k), a reduction in vOTU count was observed across both NGS and NGS + HiFi samples (Fig. 1a; Table S5). Notably, the average length of vOTUs derived from NGS + HiFi was significantly longer than those obtained solely from NGS ($P<0.0001$, Fig. 1a; Table S5). To avoid the potential exclusion of significant data due to overly restrictive length thresholds, we opted for a minimum contig length of 10 k as the benchmark for our analysis. Under this filtering condition, the quality of vOTUs derived from NGS + HiFi was superior to that obtained from NGS alone. Specifically, the number of complete vOTUs increased by 91.6%, rising from 132 to 253 (Fig. 1b; Table S5). Additionally, the overall proportion of complete, high-quality, and medium-quality vOTUs (completeness ≥ 50) improved from 7.8% to 20.0% with NGS + HiFi compared to NGS alone (Fig. 1b; Table S5).

Following length and completeness filtration, we assembled a total of 6,922 vOTUs from the combined NGS and HiFi sequencing data. Comparative analysis revealed that 2,352 vOTUs (33.98%) are present in the RVD database, 2,149 vOTUs (31.05%) in the URPC database, and 608 vOTUs (8.78%) are found in both (Fig. S1). Consequently, a substantial proportion—3,893 vOTUs (56.24%)—identified in this study represent novel entities not catalogued in either of the established databases.

Based on the life cycle annotation, the 6,922 vOTUs comprised 4,716 lytic, 1,961 temperate, and 245 unknown vOTUs (Fig. 1c; Table S6). At the family level, the majority of vOTUs (57.87%) remained unannotated. The family *Siphoviridae* was the most dominant, constituting 23.68% of the total relative abundance, followed by *Myoviridae,* which accounted for 7.96% of the total abundance (Fig. 1c; Tables S6 and S7). The vOTUs spanned from the longest at 318,474 bp to the shortest at 10,031 bp, with an average length of $4{,}421.4 \pm 343.3$ bp per sample (Fig. 1c; Table S6). In total, 5,665 vOTUs were detected or optimized by the HiFi sequencing platform, while 1,257 vOTUs were obtained from the NGS platform (Fig. 1c; Table S6). Host annotations were available for 1,963 vOTUs, with the most prevalent hosts identified as *Firmicutes_A* (33.60%) and *Bacteroidota* (33.24%) at the phylum level (Fig. 1c; Tables S8 and S9).

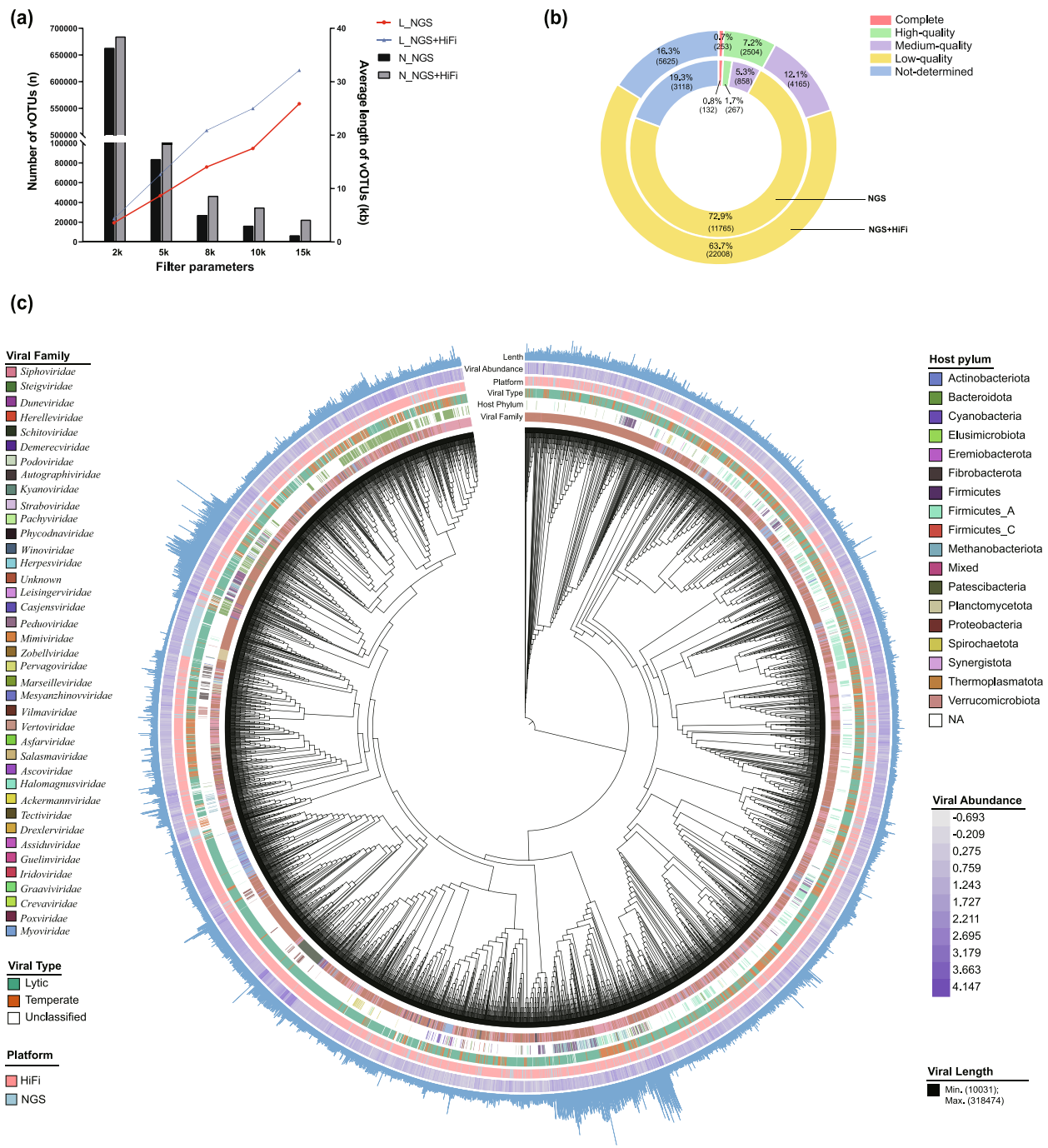

Fig. 1 Comparison of NGS and HiFi sequencing technologies and overview of DNA virome. **a** The number and average length of vOTUs under the threshold of contigs greater than 2 k, 5 k, 8 k, 10 k or 15 k, respectively. **b** The quality of vOTUs obtained by NGS and NGS + HiFi when the length of contigs greater than 10 k. **c** Overview of DNA virome in dairy cows including family taxonomy of vOTUs, life-cycle types of vOTUs, sequencing platforms, phylum taxonomy of vOTUs' host, relative abundance of vOTUs, and length of vOTUs

Structural composition of life cycle-dependent viruses in the rumen

Given the distinct life cycles of lytic and temperate viruses, the mechanisms by which these viral types exert their influence differ. Therefore, we analyzed these two categories separately in subsequent studies. Across all 30 samples, the counts ($P < 0.0001$), relative abundance ($P < 0.0001$), and Shannon index ($P = 0.0277$) of lytic

vOTUs were significantly higher than those of temperate vOTUs (Fig. 2a). At the family, genus, and species taxonomic levels, both the quantity and diversity of lytic viruses surpassed those of temperate viruses (Fig. 2b). However, there was overlap between lytic and temperate viruses at these taxonomic levels (Fig. 2b). The total abundance of lytic vOTUs constituted 82.49 ± 1.06% of the overall combined vOTU abundance, whereas temperate vOTUs made up 17.60 ± 1.06% (Tables S6 and S7). At the family level, the majority of vOTUs remained unannotated, representing 40.78% and 47.59% of the relative abundance in lytic and temperate viruses, respectively (Fig. 2c). In a comparative analysis of the 10 most abundant viral families between lytic and temperate viruses, shared families included *Siphoviridae*, *Myoviridae*, *Podoviridae*, *Phycodnaviridae*, and *Salasmaviridae*. Families unique to lytic viruses were *Schitoviridae*, *Demerecviridae*, *Straboviridae*, *Casjensviridae*, and *Herelleviridae*, while temperate virus-specific families included *Duneviridae*, *Autographiviridae*, *Winoviridae*, *Steigviridae*, and *Iridoviridae* (Fig. 2c). Except for *Phycodnaviridae*, which are eukaryotic viruses that infect large algae, the remaining viruses are bacteriophages, predominantly classified within the order *Caudovirales*, with a minority belonging to the order *Microviridae*. In the lytic viruses, the families *Siphoviridae* and *Schitoviridae* were the most prevalent, accounting for 30.35% and 23.93% of the annotated vOTUs, respectively. For temperate viruses, the *Siphoviridae* family was the most dominant, representing 67.21% of the total annotated vOTUs (Fig. 2c; Tables S6 and S7).

AMGs of life cycle-dependent viruses in the rumen

To eliminate the influence of host genomic sequences on the functional composition of the virome, we annotated AMGs on residual viral sequences after excluding provirus sequences. A total of 6,868 non-provirus vOTUs were selected for AMG prediction, and ultimately, 2,382 vOTU sequences were predicted across 56 AMG categories. Among these, 1,752 (51 categories) AMGs were associated with lytic viruses, and 589 (29 categories) AMGs with temperate viruses (Figs. S2a and 2b). The majority (79.5%) of detected vOTUs carried only one AMG, while the remaining vOTUs harbored two or more AMGs (Fig. S2e; Table S10). Specifically, two lytic vOTUs, vOTU3931 and vOTU3877, carried six AMGs (Fig. S2e; Table S10). AMGs related to carbon metabolism were the most

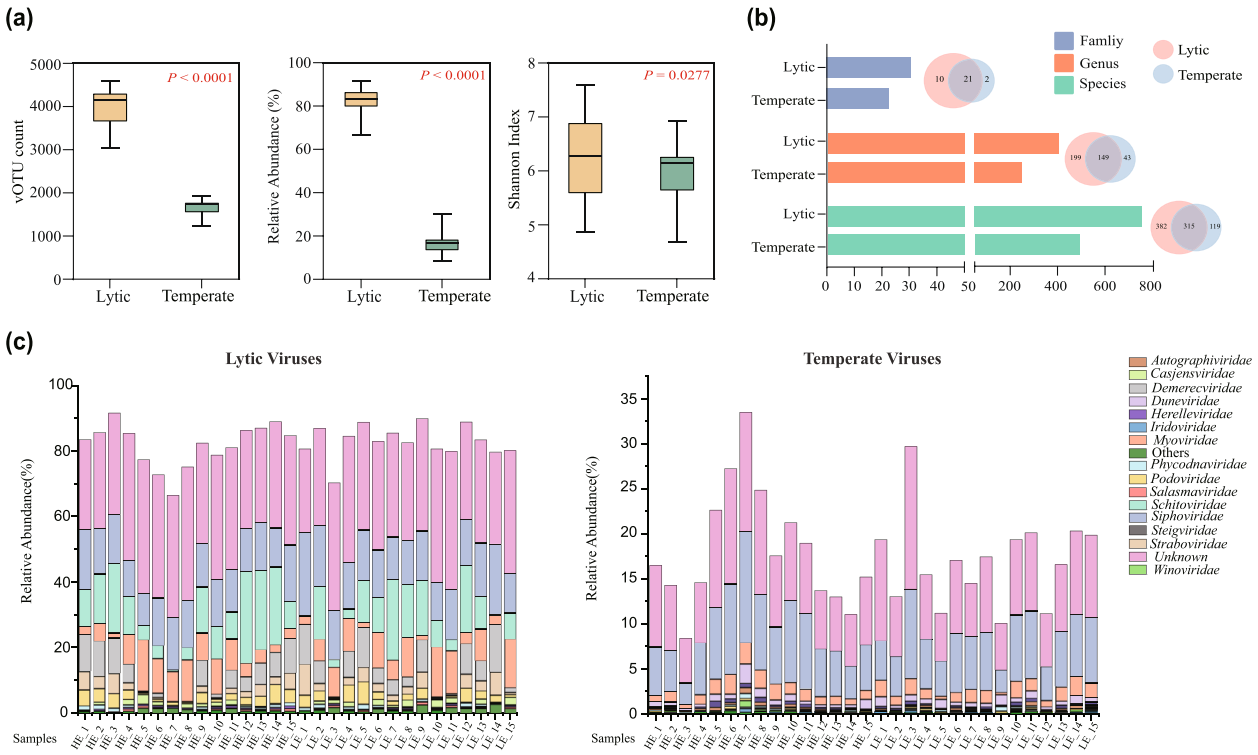

Fig. 2 Structural composition of lytic and temperate viruses in all samples. **a** The number of vOTU (left) and the α diversity of vOTU between lytic and temperate viruses (middle: richness indexes, right: shannon indexes). The significance of the differences was determined by Mann–Whitney U test. **b** The number of vOTUs and viruses at family, genus, and species levels. The bar chart represents the number of vOTUs. The red color in the pie chart represents lytic viruses, and the blue color represents temperate viruses. **c** Community compositions (family level) of lytic and temperate viruses, only shows the most 10 abundant viruses of the two types of viruses

numerous, comprising 29 categories and representing the largest group. The miscellaneous (MISC) metabolism category was the second largest, encompassing 14 AMG categories (Fig. 3a and S2d; Table S10). However, the count of AMGs involved in MISC metabolism was the highest (Table S10). The most prevalent AMG in both lytic and temperate viruses was *dut*, involved in nucleotide metabolism, encoding dUTP pyrophosphatase, which hydrolyzes dUTP to dUMP and pyrophosphate (Figs. S2c and 2d; Table S10). Additionally, other genes such as *nrdD*, *metK*, and *DNMT1* were abundant in both lytic and temperate viruses. Notably, a lytic virus-specific AMG, *cysH*, was also highly abundant, encoding an adenosine phosphate sulfate reductase protein involved in the synthesis of sulfite from sulfate. Among the top 10 most abundant AMGs for both lytic and temperate viruses, *dut*, *nrdD*, *metK*, *DNMT1*, *GT2*, *GT8*, and *GT32* were present, with *dut* being the most numerous in both categories (Fig. S2d). It was also observed that carbohydrate-active enzymes (CAZymes) predominated among the top 10 AMGs in temperate viruses, whereas lytic virus AMGs were more evenly distributed across MISC, organic nitrogen, and carbon utilization categories (Fig. S2d). To clearly illustrate the metabolic pathways of the host affected by AMGs, we divided the pathways into nine sub-categories (Fig. 3b; Table S10). Exclusive to lytic viruses were 25 AMG categories, while three categories were unique to temperate viruses (Figs. 3b and S2a). For instance, lytic-specific *fliC* and *fliD* are involved in flagellar structure metabolism, whereas temperate-specific *GH28* and *GT25* belong to CAZymes (Fig. 3b; Table S10). Additionally, 26 AMG categories were shared between the two viral types (Figs. 3b and S2a; Table S10).

Host prediction and correlation of life cycle-dependent viruses

A total of 2,232 virus-host linkages were annotated, comprising 1,377 lytic vOTU-hMAG (host metagenome-assembled genome) and 780 temperate vOTU-hMAG linkages (Table S8). Among these, 2,067 linkages involved bacterial MAGs, and 165 involved archaeal MAGs. The hosts of lytic vOTUs were categorized into 458 MAGs, whereas those of temperate vOTUs were associated with 366 MAGs (Table S8). The majority of vOTUs (88.95%) were linked to a single host, with co-hosted vOTUs typically originating from the same family (Fig. S3, Table S8). Notably, vOTU4641, vOTU6113, and vOTU6884 exhibited the highest host diversity, each associated with five distinct hosts (Table S8). Furthermore, 32 vOTUs were identified with multiple hosts at the family level; for instance, vOTU6884 was linked to five hosts spanning three different families: *UBA66*, *Bacteroidaceae*, and *Paludibacteraceae* (Table S8). Intriguingly, CowSGB-5648, an archaeal MAG affiliated with *Methanobrevibacter*, the predominant genus of methanogens, was predicted to host 140 vOTUs, of which 122 (87.14%) were lytic vOTUs and 17 (12.14%) were temperate vOTUs (Table S8).

Correlation analysis revealed 4 negatively correlated pairs and 578 positively correlated pairs between lytic vOTUs and their hosts at the MAG level (Spearman, $P<0.05$, $|r|\geq 0.5$) (Table S11). Additionally, 294 pairs demonstrated positive correlations between temperate vOTUs and their hosts (Spearman, $P<0.05$, $|r|\geq 0.5$) (Table S11). The hosts of lytic viruses were distributed across 15 different phyla, while those of temperate viruses were associated with 11 different phyla. Predominantly, the hosts for both viral types were affiliated with Firmicutes_A (comprising 32.2% of lytic and 38.9% temperate viral hosts) and Bacteroidota (comprising 24.4% of lytic and 45.0% temperate viral hosts) (Fig. 4). This distribution is consistent with the two bacterial phyla exhibiting the highest relative abundance (Figs. S4, Table S9).

Differences in rumen viruses between HE and LE groups

The structural composition of the rumen viral community was meticulously examined to identify any variations between the HE and LE groups. No significant differences were observed in the Richness ($P=0.2854$), Shannon ($P=0.9349$), and Simpson ($P=0.6827$) indices of total vOTUs between the HE and LE groups (Fig. 5a). This pattern held true for both lytic (Fig. S5a) and temperate vOTUs (Fig. S5c). Principal coordinates analysis further revealed no disparities in vOTU counts and the distribution of vOTU relative abundance between the two groups (Fig. 5b, Figs. S5b and 5d). Moreover, the total relative abundance of lytic and temperate viruses did not differ significantly between the two groups (Fig. 5c). In the correlation analysis between vOTU relative abundance and FE-related phenotypes (DMI, milk yield, milk protein content, and milk fat content), 23 lytic (19 positive, 4 negative) and 11 temperate (10 positive, 1 negative) viruses exhibited significant correlations with ECM/DMI (Spearman, $P<0.05$, $|r|\geq 0.5$) (Fig. 5d).

Using STAMP software for comparative analysis, significant differences in viral abundance were observed between the HE and LE groups. At the family level, *Drexlerviridae* ($P=0.0255$) exhibited significantly lower abundance in the HE group among lytic viruses (Fig. 5e). Conversely, *Leisingerviridae* ($P=0.0396$) exhibited higher abundance in the HE group among temperate viruses (Fig. 5e). At the genus level, two lytic viral genera were more abundant in the HE group, whereas five were more abundant in the LE group; for temperate viral genera, four exhibited higher abundance in the HE group (Fig. 5e). At the species level,

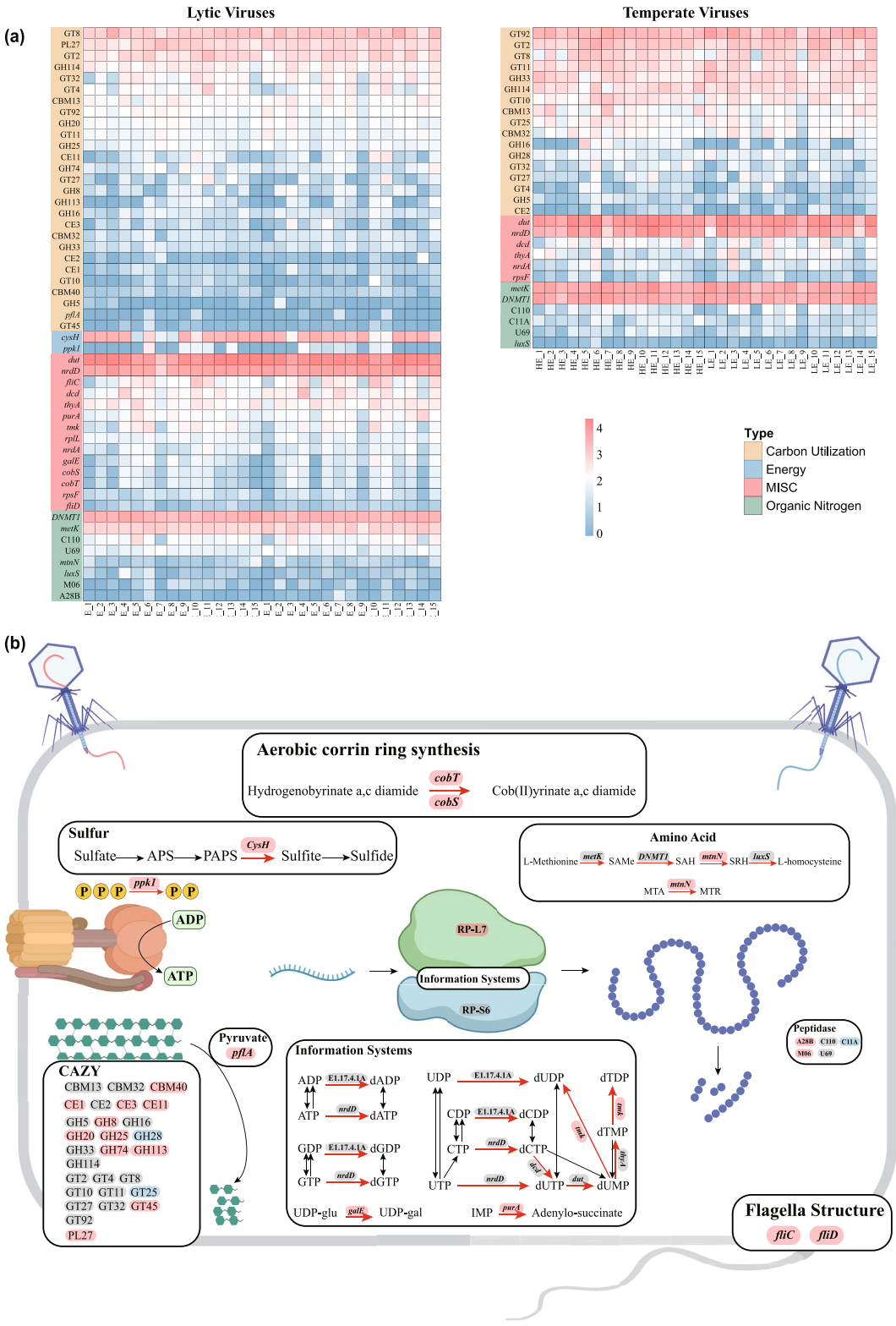

Fig. 3 Functional composition of rumen viruses and AMGs involved in virus-host interactions. **a** The relative abundance of AMGs in lytic and temperate viruses. **b** AMGs involved in virus-host interactions of 9 metabolism pathways. Red markers: AMGs specific to lytic viruses; Blue markers: AMGs specific to temperate viruses; Grey markers: AMGs both in lytic and temperate viruses. AMG: Auxiliary Metabolic Gene

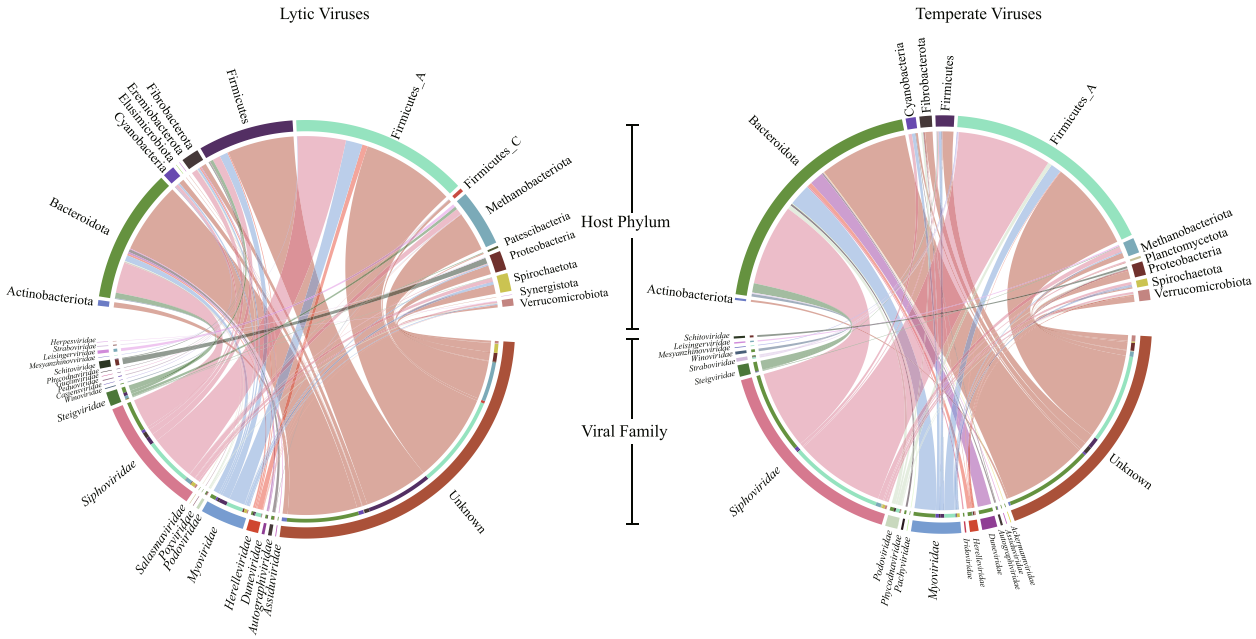

Fig. 4 Correspondences of viruses and their host. A chord diagram linking viruses to their hosts, with viruses as the starting point and hosts as the endpoints. The upper half of the figure displays hosts at the phylum level, while the lower half shows viruses at the family level. In either outer ring of the two parts, each color represents a virus/host, and the segment proportion represents the percentage of quantity within that half. The color of inner ring in the virus part represents the composition of their hosts

six lytic viral species were more abundant in the HE group, while eight were more abundant in the LE group; for temperate viruses, seven species were more abundant in the HE group, and one in the LE group (Fig. 5e). At the vOTU level, 66 lytic vOTUs exhibited higher abundance in the HE group compared to 29 in the LE group. For temperate vOTUs, 36 were more abundant in the HE group, and 11 in the LE group (Table S12).

We subsequently conducted a differential analysis of AMGs within the HE and LE groups. The findings revealed no significant differences between the two groups, regardless of whether the analysis pertained to total AMGs or to specific subtypes, such as lytic AMG and temperate AMG (Tables S13-S15). Pathway enrichment analysis of AMGs similarly identified no statistically significant discrepancies (Fig. S6, Tables S16-S18).

Among the 142 vOTUs exhibiting differential abundance between the HE and LE groups, a subset of 29 (16 lytic and 12 temperate) vOTUs were linked to 28 hMAGs, forming 38 vOTU-hMAG linkages (Table S19). Specifically, CowSGB-12222, classified within the *Bacteroidaceae* family, showed a higher abundance in the LE group ($P = 0.034$), while the remaining hMAGs did not display significant differences between the two groups ($P > 0.05$) (Table S19).

Mechanism by which rumen viruses are linked to feed efficiency

By integrating the viral life cycle, vOTU-hMAG linkages, AMGs carried by the viruses, and the impact of the host on FE-associated phenotypes (Table S20), we propose two primary mechanisms by which viruses may influence the FE of dairy cows by affecting their hosts: (1) Lytic viruses can directly lyse hosts associated with FE-related phenotypes, altering the structural and functional composition of the host microbiome in the rumen, which in turn impacts FE (Tables S8 and S20). For instance, we found that vOTU1836 and vOTU5102 can lyse CowSGB-8648, a member of the *Ruminococcaceae* family, which produces acetic acid, crucial for milk fat synthesis and FE (Tables S8 and S20). Additionally, 137 lytic vOTUs were found to lyse seven archaeal MAGs, potentially aiding in the regulation of methane emissions and energy loss, thereby improving FE. (2) Virally encoded AMGs can modulate specific metabolic pathways within the host, leading to alterations in the ruminal metabolic profile, which subsequently affects FE (Tables S11 and S20). For example, our findings indicated that temperate vOTU0897, which encodes the GT2 glycoside hydrolase, plays a pivotal role in carbohydrate degradation and utilization. Its host, CowSGB-4586, belonging to the *Lachnospiraceae* family, significantly contributes to the production of short-chain fatty acids, thereby influencing

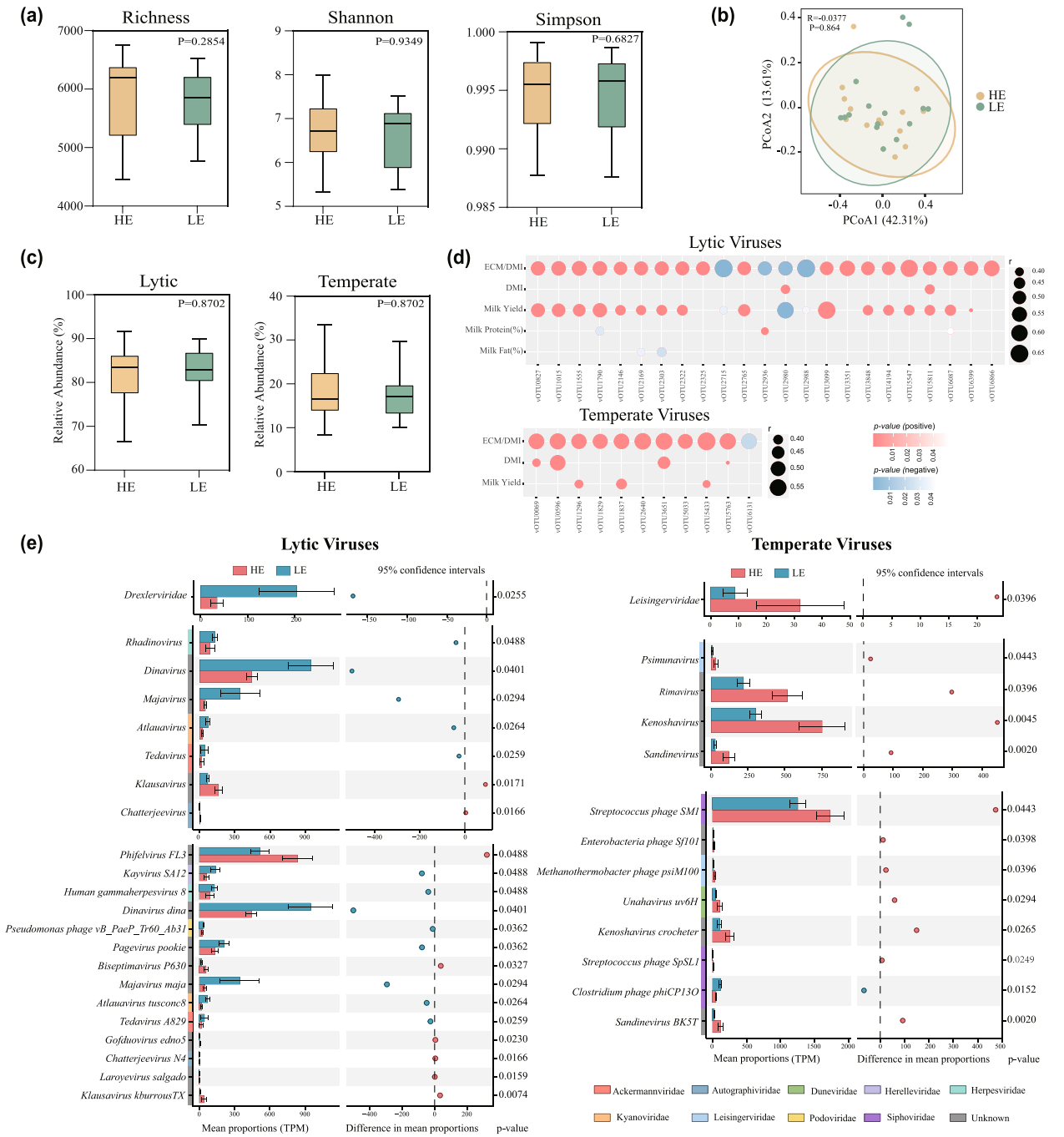

Fig. 5 Compositional differences of the rumen viruses between the HE and LE groups. **a** The α diversity of vOTUs between the HE and LE groups. The significance of the differences was determined by Mann–Whitney U test. **b** The β diversity of vOTUs between the HE and LE groups based on Bray–Curtis dissimilarity. **c** The relative abundance of lytic and temperate vOTUs between the HE and LE groups. **d** The Spearman's correlations between the relative abundance of lytic/temperate vOTUs and the phenotypes of dairy cows. **e** Significantly different vOTUs at family, genus, and species level between the HE and LE groups based on STAMP analysis (wilcox.test, 95% confidence intervals). HE: high feed efficiency; LE: low feed efficiency; STMAP: Spatial Transcriptomics Analysis Pipeline

the FE of dairy cows. To further elucidate the mechanisms discussed in this paper, we have used the 29 differentially abundant vOTUs between the HE and LE groups as a starting point and combined this information to construct Fig. 6, with the hope of providing valuable insights for future research (Fig. 6).

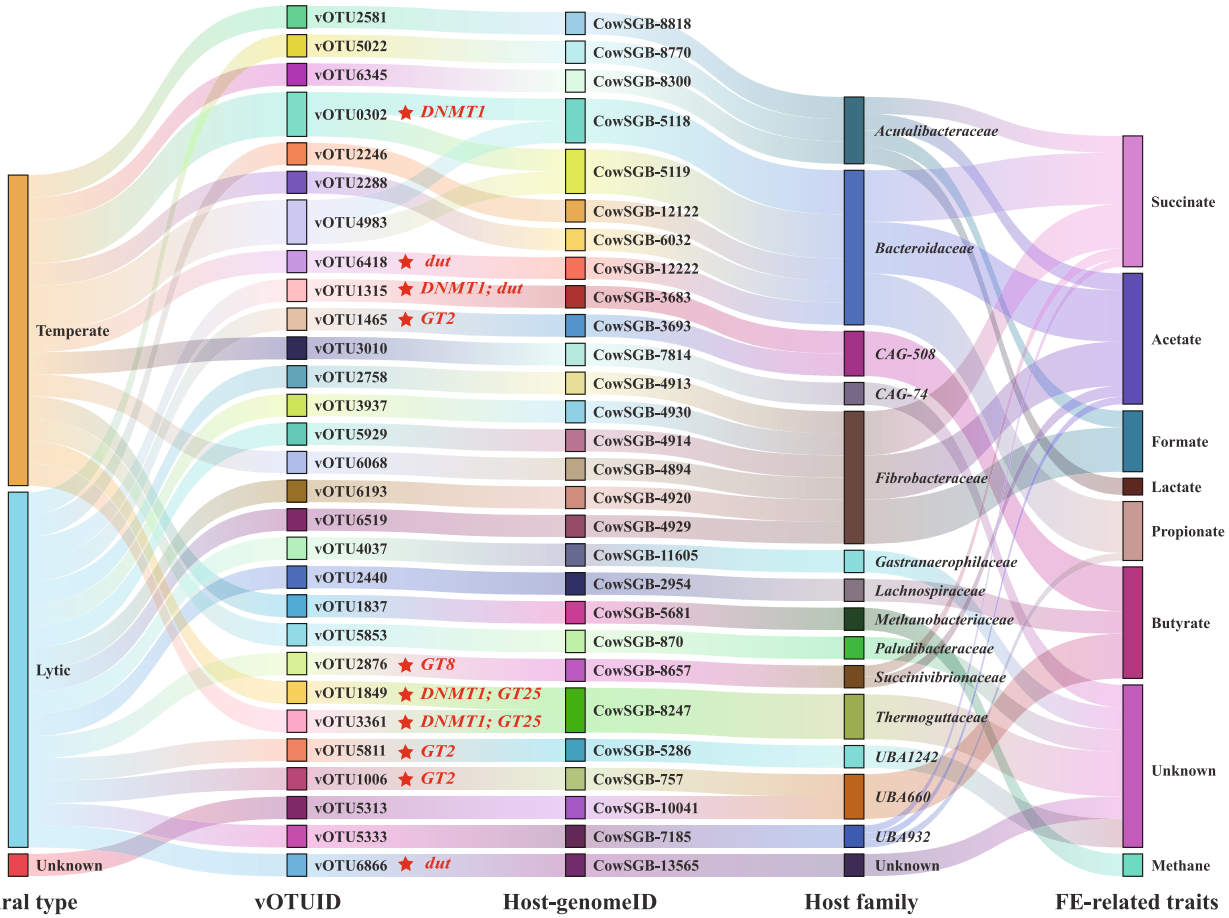

Fig. 6 Mechanism of which rumen viruses impact feed efficiency through affecting their host. The figure shows the differential vOTU between the HE and LE groups, viral type, host, FE-related traits of host, and AMGs carried by the vOTU (red star). The lytic viruses can lyse their host directly, and viruses carrying AMGs can alter the metabolism of their host, and the structural and functional composition of the rumen microbiome may be changed, thus influencing the productive traits of dairy cows. HE: high Feed efficiency; LE: low feed efficiency

Discussion

As the "dark matter" of the gut microbiome, the virome and its associated roles have attracted increasing attention. However, limitations in sequencing depth and computational tools have hindered the elucidation of gut viruses' linkages to human and animal phenotypes. In this study, we integrated deep NGS (5 times the conventional sequencing depth) with HiFi-based third-generation long-read sequencing to achieve a comprehensive understanding of the structural and functional composition of the rumen virome in dairy cows. By conducting differential and correlation analyses, we identified rumen viruses associated with FE. Additionally, based on predicted hosts and annotated AMGs of targeted viruses, this study offered novel insights into the potential mechanisms by which different types of viruses may influence FE in dairy cows.

To date, a limited number of studies have explored the rumen virome/phageome, typically with an average sequencing depth per sample not exceeding 10 Gb [30–34, 42, 43]. However, our research, with an average depth exceeding 30 Gb per sample, attempted to mitigate the constraints of previous studies. This increased depth may provide a more refined perspective on the viral populations within the rumen, considering the small genome size, high diversity, and low abundance of these viruses [13]. Employing a hybrid assembly strategy, we integrated NGS and HiFi reads to leverage the collective advantages of these sequencing methodologies. Although HiFi sequencing was conducted on only one sample (two cells), our results suggest that the advantages observed in the long reads from this single sample can be extrapolated to 30 samples. Studies comparing hybrid assembly with short-read sequencing and long-read sequencing have revealed that an integrated approach using both short and long reads holds the potential to produce high-quality assemblies while reducing costs [44, 45]. Consistent with our findings, Eisenhofer et al. [46] integrated

NGS ($n=22$) and HiFi ($n=2$) sequencing and found that, while the hybrid assembly method was not flawless, it could achieve the highest assembly coverage rate. Additionally, it performed relatively well in terms of assembly contiguity and length, as well as the number, contiguity, and completeness of MAGs. Therefore, this hybrid assembly that yields high-quality data has provided assistance in saving time and cost.

The establishment of the RVD [33] and URPC [34] has enabled extensive analysis of the diversity, virus-host linkages, and potential roles of the rumen virome within the ecosystem. However, upon comparison, the vOTUs obtained in our study exhibited over half (56.24%) novelty relative to these databases, which not only implied the high diversity of viruses in the rumen but also highlighted the quality enhancement of vOTUs achieved through our high sequencing depth and the NGS + HiFi assembly approach. Similar to other studies, a significant portion of the viruses in our research remains unannotated, indicating that many "unresolved mysteries" still surround rumen viruses. Interestingly, at the family level, *Siphoviridae* consistently emerged as the most prevalent, yet the secondary and tertiary rankings differed among the three studies. In our study, the ranking was *Siphoviridae* (34.9% of known viruses), *Schitoviridae* (19.0%), and *Myoviridae* (14.5%); in the RVD, it was *Siphoviridae* (79.6%), *Myoviridae* (14.3%), and *Podoviridae* (4.5%); and in the URPC, it was *Siphoviridae* (34.0%), *Podoviridae* (13.0%), and *Mimiviridae* (11.2%). Although this discrepancy is partly attributable to the methods of selecting representative viruses and the tools used for annotation, it also reflects the specificity of each study, suggesting that rumen viral databases still require substantial data to refine their comprehensiveness. Moreover, since RVD and URPC encompass a variety of ruminant species, including Holstein cows and 12 other species, these variations may reflect specific viral populations or ecological niches within the rumen. Therefore, our data contribute to a more accurate description of the composition of DNA viruses in the rumen of Holstein cows. Given the different life cycles of lytic and temperate viruses and their respective functions [15], the structural and functional composition of lytic and temperate viruses should be investigated separately, which has been largely ignored in other studies. The presence of unique viral families, exemplified by the lytic-specific *Salasmaviridae* and the temperate-specific *Iridoviridae*, underscores the necessity for distinct investigations into lytic and temperate viruses, potentially offering a more nuanced understanding of their ecological roles.

Viruses establish connections with animal phenotypes through initial and direct interactions with their hosts, with AMGs serving as a significant conduit [30, 44]. Our results revealed a total of 54 types of annotated AMGs, with 28 types being CAZyme genes. A previous study reported an abundance of CAZyme modules within 186 rumen bacterial genomes [47], suggesting a long-term co-evolutionary process wherein rumen viruses have assimilated essential genes from their microbial hosts, thereby assuming supportive roles in fibrolytic pathways. The most abundant AMG identified in our study was *dut*, which encodes dUTP pyrophosphatase and can catalyze the hydrolysis of dUTP into dUMP and PPi, playing a crucial role in preventing erroneous insertions during DNA synthesis and maintaining nucleotide balance [48]. Additional highly abundant AMGs, such as *DNMT1*, *metK*, and *nrdD*, also contributes to DNA synthesis and repair [49–51], reflecting the importance of nucleic acid metabolism-related genes in rumen viruses in dairy cows. However, a significant discrepancy existed in the abundance and diversity of annotated AMGs between the RVD and our study, which may be attributed to the stringent criteria RVD employed for AMG screening [33]. A disparity was observed in the specificity of AMGs between lytic (25 categories) and temperate (3 categories) viruses, indicating a higher diversity and involvement in metabolic pathways for the lytic viruses, aligning with findings from previous environmental microbiome studies [52]. Lytic viruses necessitate specific AMGs to hijack the host's metabolic machinery for their replication, a strategy not employed by temperate viruses [52, 53]. Our findings indicated that lytic viruses possess three distinct categories of AMGs in the "information system," with roles in processes such as DNA synthesis, modification, and glycosylation. For example, *tmk* generates dUDP and dTDP, potentially involved in DNA synthesis and nucleic acid modification [54]. For temperate viruses, the AMGs implicated in carbon metabolism also exhibited elevated abundance and frequency. Given the requisite integration of temperate viruses into the host genome, it was posited that temperate viruses harboring such AMGs may augment the host's fermentative capacity, thereby conferring communal benefits and, in turn, reciprocal advantages to the viruses themselves [55].

There were no significant differences in α and β diversity of vOTUs, as well as AMGs, indicating that healthy cows within the same dairy farm may share a common viral source and exhibit a relatively stable composition of rumen viruses. However, the differentially abundant viruses, serving as pivotal "biomarkers", offered an avenue for investigating the correlation between rumen viruses and FE in dairy cows. Interestingly, at the genus and species level, most differential lytic viruses showed lower relative abundance in HE groups, whereas differential temperate viruses were more numerous in HE animals. To explain this phenomenon, we speculated that lytic

viruses may replicate within and lyse bacteria/archaea positively associated with production traits, leading to an increase in their own abundance and a decrease in FE [56]. On the other hand, temperate viruses may play beneficial roles, such as carrying AMGs that enhance host competitiveness, thereby increasing in abundance with the host. However, these hypotheses still require detailed analysis of virus-host relationships and dynamic experimental validation [57].

To elucidate the mechanisms by which viruses affect FE, we proposed two general pathways: (1) Lytic viruses directly lyse FE-associated hosts, modifying rumen composition and function, thereby impacting FE; (2) Virus-encoded AMGs alter host metabolism, affecting the ruminal metabolic profile and thus FE. Most phenotypic expressions are related to metabolism and can exert direct or indirect effects on FE. Several studies have verified correlations between ruminal metabolites (e.g., VFAs) and FE in cattle [58–60]. We established associations between viruses (and AMGs), hosts, and phenotypes of dairy cows. For example, as an important source of energy for cows [61], the concentration of acetate is higher in the rumen of high-FE cows [59], and *Lachnospiraceae* and *Ruminococcaceae* are two acetate producers. On the one hand, as a temperate virus, vOTU0897 carrying GT2 may enhance the function of acetate production in *Lachnospiraceae*, thereby improving FE. On the other hand, lytic viruses such as vOTU1836 and vOTU5102 may decrease acetate production by lysing Ruminococcaceae, consequently reducing FE. Therefore, it is plausible to explain how viruses impact their hosts by establishing a "virus-host-FE" pathway, which can further elucidate how viruses affect FE. Additionally, targeting methanogens, we identified 142 lytic vOTU-hMAG linkages, all of which have the potential to reduce methane production. A total of 224 AMGs are carried by viruses infecting methanogens, with 189 belonging to "information systems." Further investigation is needed to explore interactions between viruses carrying such AMGs, which are more conducive to viral replication and methanogenic archaea.

Although our study combined high-depth NGS and HiFi sequencing to gain a more detailed and accurate understanding of the structural and functional composition of the rumen virome in dairy cows, we did not investigate RNA viruses, which may have important functions [62]. Future research should address this gap to broaden our functional insights into ruminal viruses. Currently, many published virome studies have conducted virus filtration and enrichment [30, 52, 63, 64]. However, we did not perform the aforementioned procedures and proceeded directly to metagenomic sequencing of the rumen fluid samples. Additionally, to minimize bias caused by a low proportion of viral reads in whole DNA sequencing, we selected a deep sequencing depth. Since neither of these methods seems to avoid omitting information from temperate viruses, future research should explore the differences and specificities of these two methods. In this study, we selected the differential viruses between the HE and LE groups as "biomarkers" and then linked them with their hosts and FE of dairy cows. Taking a different perspective, the hosts can also serve as "biomarkers." Many differential bacteria/archaea between the HE and LE groups have already been reported [2, 5, 65]; for example, *Selenomonas*, which is positively related to FE, has seven vOTUs in our results. By integrating the virus-host linkages, AMGs, and host-trait relationships, both types of "biomarkers" can provide insights for elucidating the mechanism of "virus-host-FE" in our study. Moreover, explicit experimental validation is required to confirm the proposed mechanisms.

Conclusion

Research on the viruses of the bovine rumen remains limited. By integrating NGS and HiFi sequencing technologies, we conducted an in-depth analysis of viral communities in 30 rumen fluid samples. Our study preliminarily established a comprehensive rumen DNA virome profile for dairy cows, encompassing viral structural and functional compositions, virus-host interactions, and AMGs, thereby contributing to the enhancement of the rumen virome database. We also proposed mechanistic pathways through which viruses may influence feed efficiency, either by lysing host cells or by inducing metabolic alterations in the host. Specific explanations were provided based on differential viral populations identified between HE and LE groups. This virus-centric approach offers novel insights into the regulatory mechanisms that link viral activity to microbial dynamics and ultimately to animal phenotypes.

Method

Animals and samples

All procedures involving animals were approved by the Animal Use and Health Committee of Zhejiang University (Hangzhou, China, No. 12410). A total of 53 multiparous mid-lactating Holstein cows, with an average body weight of 634 ± 85 kg and 153 ± 20 days in milk, were selected for a 57-day study (including a 7-day adaptation period). During this time, the cows were housed in a free-stall barn with access to a total mixed ration and were fed three times daily (06:30, 14:30, and 21:30) with a corn-based high-grain diet (Table S21). Body weights were recorded immediately after the morning milking every week, and daily dry matter intake (DMI) was measured using an automatic feed system (Zhenghong

Co., Shanghai, China). Milk yield was recorded at three milking times (06:00, 14:00, and 21:00), with subsamples stored at 4 °C and analyzed for milk composition using infrared spectroscopy. Rumen fluid samples were collected via an oral stomach tube before morning feeding on day 50 and stored at −80 °C until analysis. From the days 8 to 56, compute daily ECM values and derive the mean; the formula for ECM is ECM = (0.3246 × kg of milk) + (13.86 × kg of milk fat) + (7.04 × kg of milk protein). Concurrently, calculate the mean DMI over the same period. FE was calculated as ECM/DMI, and 15 high-efficiency (HE, ECM/DMI = 1.593 ± 0.063) and 15 low-efficiency (LE, ECM/DMI = 1.374 ± 0.071) cows were selected for further analysis.

DNA extraction, library preparation, and next-generation sequencing

Rumen fluid samples were not filtered before DNA extraction to retain all free viruses and prophages. DNA was extracted from 1.5 g of rumen fluid using the E.Z.N.A.® Stool DNA Kit (Omega Bio-tek, Norcross, GA, USA) following the manufacturer's protocols. DNA purity and concentration were confirmed using NanoDrop ONE and Qubit 4.0, respectively, and all samples met the required standards for further analysis (Table S22). NGS libraries were prepared using the TruSeq™ Nano DNA Sample Preparation Kit (Illumina, San Diego, CA) with 1 μg of total DNA. DNA end repair, A-base addition, and Illumina-indexed adaptor ligation were performed according to the manufacturer's protocol. Libraries were size-selected for DNA fragments of ~ 400 bp using 2% Low Range Ultra Agarose, followed by PCR amplification with Phusion DNA polymerase (New England Biolabs, USA) for 15 cycles. Sequencing was performed on the Illumina NovaSeq 6000 platform (150 bp × 2, Shanghai Biozeron Biotechnology Co., Ltd, Shanghai, China).

HiFi sequencing

A total of 5 μg of DNA from the HE_6 sample was used to prepare a SMRTbell library using the PacBio SMRTbell Prep Kit 3.0 (Pacific Biosciences, CA, USA, Part Number: 102–182–700) according to the manufacturer's recommendations. Damaged double-stranded DNA was repaired using the New England BioLabs PreCR® Repair Mix Kit before library preparation. Repaired DNA was size-selected using the BluePippin system (Sage Science, MA, USA) to obtain molecules larger than 3 kb. The SMRTbell library was sequenced with v3 chemistry on a PacBio Sequel IIe instrument (Pacific Biosciences, CA, USA) using SMRT 8 M cells (Part Number: 101–389-001). HiFi reads were generated using the 'ccs' module (parameters: –min-length 500 –min-passes 3 –min-rq 0.99) within the SMRT Link v10.0 package (Pacific Biosciences, CA, USA).

NGS and NGS + HiFi assembly processes

For NGS, raw sequence reads were quality trimmed using Trimmomatic (v0.36) [66] to remove adaptor contaminants and low-quality reads. Reads that passed quality control were mapped against the bovine genome using the BWA mem algorithm [67], with parameters "-M -k 32 -t 16". Residual high-quality reads, devoid of host genomic contamination, were designated as clean reads for downstream analysis. Clean reads were assembled into contigs for each sample using MegaHit [68] (v1.1.1–2-g02102e1) with "–min-contig-len 500" parameters.

For NGS + HiFi, HiFi long reads were initially corrected using MaSuRCA (v4.0.3) [69], followed by assembly of the NGS short reads and HiFi long reads using a hybrid approach with metaSPAdes (v3.15.3) [70] and Operons (v2.11-r797) [71].

vOTU identification and taxonomic classification

For enhanced accuracy, viral contigs from metagenomic assemblies were identified using VirSorter2 (v2.2.3, max_score ≥ 0.7) [72] and DeepVirFinder (v1.0, Score ≥ 0.7, p-value ≤ 0.05) [73]. Viral contigs longer than 10 k were manually screened for subsequent analysis. CheckV v0.7.0 [74] was employed to remove non-viral sequences during VirSorter2 analysis. All potential viral contigs were further validated using VIBRANT v1.2.1 [75]. Identified viral contigs were clustered using MUMmer software [76] (95% ANI, ≥ 85% coverage), and the longest representative contig within each cluster was designated as a vOTU. To prevent information loss from overly large filter lengths, 10k was chosen as the filter parameter, resulting in 6,922 vOTUs for further analysis. Virus lysis was detected by VIBRANT (v1.2.1) and proviruses were detected by CheckV (v0.7.0). Viral taxonomic annotation was performed using tBLASTX [77] based on the NCBI virus classification list (e-value ≤ 1e-5, as of November 2022). Given the significant genomic differences between vOTUs, a similarity matrix at the protein level was generated by ViPTree (v1.1.2) [78], leading to the construction of a phylogenetic tree.

Comparisons of the viral genome in this article with the RVD and URPC databases

For the assessment of vOTU novelty, a comparison was conducted against 397,180 RVD and 64,922 URPC viral genomes utilizing BLASTn with parameters -max_target_seqs 5 -evalue 1e-5 -outfmt 6. Viral genomes with ≥ 90% identity and ≥ 75% coverage were considered to be the same, with the remaining vOTUs considered novel.

Calculation of vOTU and MAG relative abundance

The relative abundances of vOTUs or MAGs in each sample were calculated as TPM (transcripts per kilobase of exon model per million mapped reads) by normalizing the read counts of each vOTU or MAG to the relative quantity in a million sequences, using in 'quant_bins.sh' script (parameters: -b bins_directory -o output_directory -a assembly_contigs -t 40) within the MetaWRAP pipeline (v1.3.2) [79].

Virus-host linkage prediction

To extend virus-host linkage predictions, we incorporated downloaded and reconstructed prokaryotic genomes [80] as a reference host database. The 6,922 vOTUs were putatively linked to 13,572 MAGs (13,412 bacterial and 160 archaeal genomes) using two in silico methods: (1) CRISPR spacers in MAGs were identified by CRT, CRISPRFinder, and CRISPRDetect [81–83]. Spacer sequences were BLASTN against viral contigs, and only matches with ≥ 30 bp alignment length and ≥ 97% nucleotide identity were considered positive virus-host matches; (2) VirHostMatcher [84] with "distance ≤ 0.2". Hosts predicted by either method were combined into the final potential host database. All virus-host linkages and the respective methods used are listed in Table S8.

AMG identification, classification, and abundance

To more accurately predict and assign levels of viral carbohydrate metabolism substrates and their contribution to geochemical cycling, DRAM-v (v1.2.0) [85] was chosen to annotate AMGs, using "default parameters, 1 ≤ score ≤ 3, AMG flags of -M and -F".

Statistical analysis

Data statistics and visualization were performed using R (v4.2.1) and GraphPad Prism 8.0. Alpha diversity between HE and LE groups was assessed using the Mann–Whitney U test. PCoA analysis clustered vOTUs from different samples based on Bray–Curtis distances. Spearman correlations were calculated to reveal relationships between vOTU relative abundance and production traits of dairy cows. Differential viruses between HE and LE groups were identified using STAMP analysis (v2.0.0) [86]. Spearman correlations were also used to determine relationships between the relative abundance of vOTUs and their corresponding hosts.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-02019-0.

Supplementary Material 1.

Acknowledgements

We acknowledge the members of the Institute of Dairy Science at Zhejiang University (Hangzhou, China) for their assistance with field sampling and data analysis.

Authors' contributions

HS supervised the project. XL and HS designed the study. XL and YT conducted bioinformatic analysis. XL wrote the initial draft of the manuscript. HC collected data. HS and JL edited and revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was supported by grants from the National Natural Science Foundation of China (32322077, Beijing).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Animal care and experimental procedures were approved by the Animal Care Committee of Zhejiang University (Hangzhou, China), and were under the university's guidelines for animal research.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 April 2024 Accepted: 19 December 2024
Published online: 16 January 2025

References

1. Liu X, Tang Y, Wu J, Liu J-X, Sun H-Z. Feedomics provides bidirectional omics strategies between genetics and nutrition for improved production in cattle. Anim Nutr. 2022;9:314–9.
2. Xue M-Y, Xie Y-Y, Zhong Y, Ma X-J, Sun H-Z, Liu J-X. Integrated meta-omics reveals new ruminal microbial features associated with feed efficiency in dairy cattle. Microbiome. 2022;10:32.
3. Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. PNAS. 2014;111:10714–9.
4. Tan RSG, Zhou M, Li F, Guan LL. Identifying active rumen epithelial associated bacteria and archaea in beef cattle divergent in feed efficiency using total RNA-seq. Curr Res Microb Sci. 2021;2:100064.
5. Xie Y, Sun H, Xue M, Liu J. Metagenomics reveals differences in microbial composition and metabolic functions in the rumen of dairy cows with different residual feed intake. Anim Microbiome. 2022;4:19.
6. Lopes DRG, La Reau AJ, Duarte MDS, Detmann E, Bento CBP, Mercadante MEZ, et al. The bacterial and fungal microbiota of Nelore steers is dynamic across the gastrointestinal tract and its fecal-associated microbiota is correlated to feed efficiency. Front Microbiol. 2019;10:1263.
7. Zhang Y, Li F, Chen Y, Wu H, Meng Q, Guan LL. Metatranscriptomic profiling reveals the effect of breed on active rumen eukaryotic composition in beef cattle with varied feed efficiency. Front Microbiol. 2020;11:367.
8. Cobián Güemes AG, Youle M, Cantú VA, Felts B, Nulton J, Rohwer F. Viruses as winners in the game of life. Annu Rev Virol. 2016;3:197–214.
9. Paynter MJB, Ewert DL, Chalupa W. Some morphological types of bacteriophages in bovine rumen contents. Appl Microbiol. 1969;18:942–3.
10. Klieve AV, Bauchop T. Morphological diversity of ruminal bacteriophages from sheep and cattle. Appl Environ Microbiol. 1988;54:1637–41.
11. Klieve AV, Swain RA. Estimation of Ruminal Bacteriophage Numbers by Pulsed- Field Gel Etectrophoresis and Laser Densitometry. Appl Environ Microbiol. 1993;59:2299–303.

12. Sato Y, Takebe H, Tominaga K, Yasuda J, Kumagai H, Hirooka H, et al. A rumen virosphere with implications of contribution to fermentation and methane production, and endemism in cattle breeds and individuals. Appl Environ Microbiol. 2024;90:e01581-23. Cann I, editor.
13. Lobo RR, Faciola AP. Ruminal phages – a review. Front Microbiol. 2021;12:763416.
14. Yan M, Yu Z. Viruses contribute to microbial diversification in the rumen ecosystem and are associated with certain animal production traits. Microbiome. 2024;12:82.
15. Salmond GPC, Fineran PC. A century of the phage: past, present and future. Nat Rev Microbiol. 2015;13:777–86.
16. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J. 2017;11:1511–20.
17. Paul JH. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? ISME J. 2008;2:579–89.
18. Yuan L, Hensley C, Mahsoub HM, Ramesh AK, Zhou P. Microbiota in viral infection and disease in humans and farm animals. Prog Mol Biol Transl Sci. 2020;171:15–60.
19. Kieft K, Zhou Z, Anderson RE, Buchan A, Campbell BJ, Hallam SJ, et al. Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. Nat Commun. 2021;12:3503.
20. Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nat Rev Microbiol. 2015;13:641–50.
21. Payet JP, Suttle CA. To kill or not to kill: The balance between lytic and lysogenic viral infection is driven by trophic status. Limnol & Oceanogr. 2013;58:465–74.
22. Rohwer F, Thurber RV. Viruses manipulate the marine environment. Nature. 2009;459:207–12.
23. Crummett LT, Puxty RJ, Weihe C, Marston MF, Martiny JBH. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. Virology. 2016;499:219–29.
24. Weynberg KD, Laffy PW, Wood-Charlson EM, Turaev D, Rattei T, Webster NS, et al. Coral-associated viral communities show high levels of diversity and host auxiliary functions. PeerJ. 2017;5:e4054.
25. Brailsford MD, Hartman PA. Characterization of *Streptococcus durans* bacteriophages. Can J Microbiol. 1968;14:397–402.
26. Hoogenraad NJ, Holmes I, Millis NF. Bacteriophages in Rumen Contents of Sheep. J Gen Virol. 1967;1:575–6.
27. Gilbert RA, Townsend EM, Crew KS, Hitch TCA, Friedersdorff JCA, Creevey CJ, et al. Rumen virus populations: technological advances enhancing current understanding. Front Microbiol. 2020;11:450.
28. Lockington RA, Attwood GT, Brooker JD. Isolation and characterization of a temperate bacteriophage from the ruminal anaerobe Selenomonas ruminantium. Appl Environ Microbiol. 1988;54:1575–80.
29. Styriak I, Kmet V, Spanova A. Isolation and characterization of two rumen Streptococcus bovis bacteriophages. Microbiologica. 1989;12:317–22.
30. Namonyo S, Wagacha M, Maina S, Wambua L, Agaba M. A metagenomic study of the rumen virome in domestic caprids. Arch Virol. 2018;163:3415–9.
31. Anderson CL, Sullivan MB, Fernando SC. Dietary energy drives the dynamic response of bovine rumen viral communities. Microbiome. 2017;5:155.
32. Solden LM, Naas AE, Roux S, Daly RA, Collins WB, Nicora CD, et al. Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. Nat Microbiol. 2018;3:1274–84.
33. Yan M, Pratama AA, Somasundaram S, Li Z, Jiang Y, Sullivan MB, et al. Interrogating the viral dark matter of the rumen ecosystem with a global virome database. Nat Commun. 2023;14:5254.
34. Wu Y, Gao N, Sun C, Feng T, Liu Q, Chen W-H. A compendium of ruminant gastrointestinal phage genomes revealed a higher proportion of lytic phages than in any other environments. Microbiome. 2024;12:69.
35. Vogl I, Eck SH, Benet-Pagès A, Greif PA, Hirv K, Kotschote S, et al. Diagnostic applications of next generation sequencing: working towards quality standards/diagnostische anwendung von next generation sequencing: Auf dem Weg zu Qualitätsstandards. LaboratoriumsMedizin. 2012;36:227–39.
36. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. Nat Biotechnol. 2008;26:1146–53.
37. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8.
38. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol. 2012;30:693–700.
39. Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. Nat Commun. 2021;12:60.
40. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol. 2019;37:1155–62.
41. Bickhart DM. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. Nat Biotechnol. 2022;40:711-19.
42. Ross EM, Petrovski S, Moate PJ, Hayes BJ. Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. BMC Microbiol. 2013;13:242.
43. Yutin N, Kapitonov VV, Koonin EV. A new family of hybrid virophages from an animal gut metagenome. Biol Direct. 2015;10:19.
44. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. HYBRID SPA DES : an algorithm for hybrid assembly of short and long reads. Bioinformatics. 2016;32:1009–15.
45. Tao Y, Xun F, Zhao C, Mao Z, Li B, Xing P, et al. Improved assembly of metagenome-assembled genomes and viruses in Tibetan saline lake sediment by HiFi Metagenomic Sequencing. Liu J, editor. Microbiol Spectr. 2023;11:e03328-22.
46. Eisenhofer R, Nesme J, Santos-Bay L, Koziol A, Sørensen SJ, Alberdi A, et al. A comparison of short-read, HiFi long-read, and hybrid strategies for genome-resolved metagenomics. Microbiol Spectr. 2024;12:e03590-23. Chen W-H, editor.
47. Xue M-Y, Wu J-J, Xie Y-Y, Zhu S-L, Zhong Y-F, Liu J-X, et al. Investigation of fiber utilization in the rumen of dairy cows based on metagenome-assembled genomes and single-cell RNA sequencing. Microbiome. 2022;10:11.
48. Harris JM, McIntosh EM, Muscat GEO. Structure/function analysis of a dUTPase: catalytic mechanism of a potential chemotherapeutic target. J Mol Biol. 1999;288:275–87.
49. Sen GL, Reuter JA, Webster DE, Zhu L, Khavari PA. DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature. 2010;463:563–7.
50. Horikawa S, Sasuga J, Shimizu K, Ozasa H, Tsukada K. Molecular cloning and nucleotide sequence of cDNA encoding the rat kidney S-adenosylmethionine synthetase. J Biol Chem. 1990;265:13683–6.
51. Logan DT, Mulliez E, Larsson K-M, Bodevin S, Atta M, Garnaud PE, et al. A metal-binding site in the catalytic subunit of anaerobic ribonucleotide reductase. PNAS. 2003;100:3826–31.
52. Luo X-Q, Wang P, Li J-L, Ahmad M, Duan L, Yin L-Z, et al. Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. Microbiome. 2022;10:190.
53. Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. Nat Microbiol. 2018;3:754–66.
54. Reichard P. Interactions between deoxyribonucleotide and DNA synthesis. Annl Rev Biochem. 1988;57:349–74.
55. Obeng N, Pratama AA, Elsas JDV. The significance of mutualistic phages for bacterial ecology and evolution. Trends Microbiol. 2016;24:440–9.
56. Voigt E, Rall BC, Chatzinotas A, Brose U, Rosenbaum B. Phage strategies facilitate bacterial coexistence under environmental variability. PeerJ. 2021;9:e12194.
57. Mukhopadhya I, Segal JP, Carding SR, Hart AL, Hold GL. The gut virome: the 'missing link' between gut bacteria and host immunity? Therap Adv Gastroenterol. 2019;12:175628481983662.
58. Na SW, Guan LL. Understanding the role of rumen epithelial host-microbe interactions in cattle feed efficiency. Animal Nutrition. 2022;10:41–53.
59. Xue MY, Sun HZ, Wu XH, Guan LL, Liu JX. Assessment of rumen bacteria in dairy cows with varied milk protein yield. J Dairy Sci. 2019;102:5031–41.
60. Xie Y, Miao C, Lu Y, Sun H, Liu J. Nitrogen metabolism and mammary gland amino acid utilization in lactating dairy cows with different residual feed intake. Anim Biosci. 2021;34:1600–6.
61. Rémond D, Ortigues I, Jouany J-P. Energy substrates for the rumen epithelium. Proc Nutr Soc. 1995;54:95–105.
62. Koonin EV, Krupovic M, Agol VI. The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution? . Microbiol Mol Biol Rev 85:https://doi.org/10.1128/mmbr.00053-21.
63. Liang G, Zhao C, Zhang H, Mattei L, Sherrill-Mix S, Bittinger K, et al. The stepwise assembly of the neonatal virome is modulated by breastfeeding. Nature. 2020;581:470–4.

64. Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. Cell Host Microbe. 2020;28:724-740.e8.
65. Xue M-Y, Sun H-Z, Wu X-H, Liu J-X, Guan LL. Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. Microbiome. 2020;8:64.
66. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
67. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013. https://doi.org/10.48550/arXiv.1303.3997.
68. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. Bioinformatics. 2015;31:1674–6.
69. Zimin AV, Puiu D, Luo M-C, Zhu T, Koren S, Marçais G, et al. Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res. 2017;27:787–92.
70. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34.
71. Bertrand D, Shaw J, Kalathiyappan M, Ng AHQ, Kumar MS, Li C, et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat Biotechnol. 2019;37:937–44.
72. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome. 2021;9:37.
73. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome. 2017;5:69.
74. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39:578–85.
75. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome. 2020;8:90.
76. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.
77. Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genet. 1993;3:266–72.
78. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Bioinformatics. 2017;33:2379–80. Valencia A, editor.
79. Uritskiy GV. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6:158.
80. Jia M, Zhu S, Xue M-Y, Chen H, Xu J, Song M, et al. Single-cell transcriptomics across 2,534 microbial species reveals functional heterogeneity in the rumen microbiome. Nat Microbiol. 2024;9:1884–98.
81. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, et al. CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics. 2007;8:209.
82. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35:W52–7.
83. Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM. CRISPRDetect: a flexible algorithm to define CRISPR arrays. BMC Genomics. 2016;17:356.
84. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res. 2017;45:39–53.
85. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res. 2020;48:8883–900.
86. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics. 2014;30:3123–4.

Publisher's Note