scientific reports

Check for updates

OPEN

Differences in prokaryotic and viral community between rumen and feces

Yoshiaki Sato[1]✉, Hajime Kumagai[2], Hiroyuki Hirooka[2] & Takashi Yoshida[3]

Ruminants harbor diverse microbial communities, including prokaryotes and viruses, across their digestive tract. Rumen viruses contribute to carbohydrate metabolism; however, their persistence and host interactions in the lower gastrointestinal tract remain unclear. In this study, we investigated the prokaryotic and viral communities in the rumen and feces of the same wethers using whole-metagenomic and virus-like particle metagenomic sequencing. For prokaryotic community analysis, we reconstructed over 300 metagenome-assembled genomes, most of which were novel. These revealed strong site specificity, with distinct prokaryotic community compositions between the rumen and feces. Virome analysis recovered more than 6,000 viral genomes, including many novel viruses. Unlike prokaryotes, several viruses were found to be shared between the rumen and feces. Auxiliary metabolic genes encoding glycoside hydrolases were identified in several rumen-associated viral genomes, whereas fecal-associated viral genomes did not harbor such genes. Host–virus interaction analysis predicted that viruses predominantly infect dominant bacterial taxa and methanogens within each gastrointestinal site, although some viruses may interact with hosts across different sites. These findings highlight the strong site specificity of the prokaryotic communities and the comparatively broader distribution of viruses within the ruminant gastrointestinal tract. These insights advance understanding of virus–prokaryote–host interactions with implications for animal productivity.

Background

Ruminants can degrade human-inedible plant materials and produce edible products such as meat and milk, contributing to global food security. This unique ability is significantly related to the gastrointestinal tract (GIT), particularly the rumen, the first chamber. The rumen plays a crucial role in anaerobic fermentation, breaking down lignocellulose with the help of microorganisms and supplying approximately 80% of the host energy requirements[1]. Several studies have demonstrated that the rumen microbiome contributes to milk[2,3] and meat production[4–6], influences feed efficiency[7,8], and drives methane emissions[9,10]. For example, rumen archaea are directly responsible for methane production[11], while specific bacterial groups affect feed efficiency by enhancing fiber degradation, increasing the production of short-chain fatty acids such as propionate[8]. Thus, numerous studies have focused on the rumen microbiota to investigate its relationship with animal production. In addition, non-rumen GIT microbiota exhibits region-specific colonization and play a role in plant feed digestion[12]. Studies have explored the relationships between the GIT microbiome and feed efficiency, daily gain, and feed intake[13]. Therefore, a comprehensive understanding of the GIT microbiome in ruminants is essential for improving animal productivity and reducing environmental impact.

Viruses infecting bacteria and archaea are present in the GIT of ruminants[14–20]. Viruses can directly regulate hosts populations through cell lysis, and some viruses harbor virus-encoded auxiliary metabolic genes (AMGs), which are expressed during infection to enhance host metabolism and viral replication[16–20]. Thus, viruses influence microbial composition and function, ultimately contributing to animal production.

Our understanding of rumen viruses has advanced the most among the different compartments of the ruminant GIT. Decades ago, researchers demonstrated that viruses, mostly tailed viruses, are abundant in the rumen and represent a critical component of its complex microbial community[21,22]. With advancements in sequencing technology, the comprehensive study of the rumen viral community ("virome") has progressed.

[1]Department of Agrobiology and Bioresources, School of Agriculture, Utsunomiya University, 350 Minemachi, Utsunomiya, Tochigi 321-8505, Japan. [2]Laboratory of Animal Husbandry Resources, Division of Applied Biosciences, Graduate School of Agriculture, Kyoto University, Kyoto, Japan. [3]Laboratory of Marine Microbiology, Division of Applied Biosciences, Graduate School of Agriculture, Kyoto University, Kyoto, Japan. ✉email: yoshis@cc.utsunomiya-u.ac.jp

As a result, rumen viruses have been found to interact with dominant rumen bacteria and archaea, such as *Prevotella*, *Ruminococcus*, and *Methanobrevibacter*, playing an important role in shaping the rumen microbial ecosystem[17–19,23,24]. Additionally, AMGs encoding glycoside hydrolases (GHs) and genes related to methane metabolism have been identified in rumen viruses[16–19,25], suggesting that rumen viruses may influence carbohydrate metabolism and methane production. Additionally, several studies demonstrated rumen viruses are significantly associated with animal production such as feed efficiency[26,27], lactation performance in dairy cattle[26], and carcass traits in beef cattle[19]. Recently, the construction of large-scale rumen DNA virus databases using whole-metagenomic sequencing[17] and virus-like particle (VLP) metagenomic sequencing[18,19] has further expanded genome-based virus studies in the rumen, paving the way for more in-depth research in this field.

In contrast, research on non-ruminal GIT viruses remains limited. For example, although several earlier studies investigated fecal viruses from ruminants, often using them to track microbial contamination of waterways through phage typing of *Bacteroides*[28], these works were not aimed at comprehensive community-level analyses. Moreover, previous studies have demonstrated that most GIT viruses are highly localized to specific GIT sites[15] and that the GIT virome varies significantly across different GIT regions[14]. However, these studies have provided only a broad overview of the GIT virome in ruminants, without detailed analyses of viruses specific to each GIT site. Additionally, these studies have relied on whole-metagenome sequencing, and no studies using VLP-metagenomic sequencing have been reported to date. Because whole-metagenomic and VLP-metagenomeicsequencing target distinct viral populations, with the former primarily detecting integrated prophages and the latter capturing free viral particles[29], VLP-based approaches are crucial for a more comprehensive characterization of the GIT virome in ruminants. Furthermore, no studies have conducted a detailed comparison of both fecal and rumen viromes within the same individual, which is critical for understanding whether these viruses persist into the lower GIT and how they interact with the host.

Herein, we investigate the rumen and fecal prokaryotic community (i.e., prokaryome) and virome of the same individual wethers using both whole-metagenomic and VLP-metagenomic sequencing. In this study, we successfully reconstructed over 300 non-redundant metagenome-assembled genomes (MAGs) in prokaryote and 6,000 non-redundant viral genomes, most of which are putatively novel. Additionally, we provide new insights into the composition and distribution of MAGs and viruses in different GIT compartments. We demonstrated that the rumen and fecal prokaryome and virome exhibit clear differences between fecal and rumen samples. Notably, rumen and fecal MAGs were highly specific to each site, whereas several viruses were shared between the two regions. Furthermore, we predicted viral-host interactions for several viruses, and some viruses may be able to infect hosts from different GIT sites.

Materials and methods

Animals, diets, experimental design, and sampling

Four ruminally cannulated Corriedale wethers at the Graduate school of Agriculture, Kyoto University, with an initial body weight (BW) of 52.9 ± 13.36 kg (mean ± standard deviation, SD) and aged from 22 to 47 months, were used. Two animals were fed a high-concentrate diet (concentrate: roughage = 70:30), while the others were offered a low-concentrate diet (concentrate: roughage = 30:70) on a dry matter (DM) basis (Supplementary Table 1). Diets were administered at 1.8% of BW on a DM basis for 14 days. The diets were evenly offered at 09:00 and 17:00 h daily. All animals were housed individually in pens with *ad libitum* access to water. On day 14, rumen fluids and feces were collected before the morning feeding. Rumen samples were filtered through four layers of cheesecloth. Rumen and fecal samples were transported on dry ice and stored at -80 °C until use.

Whole-metagenomic sequencing

The 1 mL of the rumen sample was centrifuged at 12,000×*g* at 4 °C for 15 min, and the pellet was used for DNA extraction. For fecal samples, 200 mg of feces was used. Genomic DNA was extracted using a protocol as previously reported[30]. Whole-metagenome libraries were prepared using the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina HiSeq X platform (2 × 150 bp).

VLP-metagenomic sequencing

VLP purification and DNA extraction were performed following a validated protocol[19,31]. Briefly, ~ 1 g of feces was mixed with 10 mL of 0.02 µm-filtered SM buffer and vortexed for 5 min. The 3 mL rumen and fecal suspensions were centrifuged twice at 6,000 × *g* for 30 min at 4 °C. The supernatants were sequentially filtered through 0.45 and 0.22 µm PVDF filters (Millipore, Burlington, MA, USA), mixed with an equal volume of 20% PEG-6000 (2.5 M NaCl), and incubated at 4 °C overnight. The mixture was centrifuged at 20,380 × *g* for 45 min at 4 °C, and the pellet was resuspended in 215 µL of 0.02 µm-filtered SM buffer containing 5 mg/mL lysozyme, followed by incubation at 37 °C for 1 h. The lysate was treated with a cocktail of DNases (12 U DNase I, 5 U TURBO DNase, 5 U Baseline-ZERO DNase, 25 U Benzonase) and 25 µg RNase at 37 °C for 1 h. Enzyme activity was stopped by adding EDTA (20 mM final) and heating at 70 °C for 15 min. Sodium dodecyl sulfate (final concentration of 1%) and proteinase K (final concentration, 0.5 mg/mL; TaKaRa Bio) were then added, and the mixture was incubated at 55 °C for 20 min and then at 70 °C for 5 min. DNA was extracted using phenol: chloroform: isoamyl alcohol (25:24:1) and chloroform: isoamyl alcohol (24:1), followed by precipitation with 3 M sodium acetate, isopropanol, and Dr. GenTLE precipitation carrier (Takara Bio). The DNA pellet was washed with 70% ethanol, air-dried, and resuspended in 0.02 µm-filtered MilliQ water. This process is referred to as Method 1.

To evaluate methodological differences, a second approach (referred to as Method 2) was applied to rumen samples based on a previous study[18]. VLPs were enriched using CsCl density-gradient centrifugation[18,32], and DNA was extracted using the xanthogenate method[33]. DNA from both methods was used for library preparation and sequencing as described above.

Analysis of MAGs

Low-quality reads and adapters in whole-metagenomic sequencing were removed using Trimmomatic v0.39[34] with the following options: ILLUMINACLIP:2:30:10, SLIDINGWINDOW:10:15, and MINLEN:50. Host contamination was filtered out by aligning trimmed reads to the sheep reference genome (GCF_016772045) with BWA-MEM with default parameters[35], retaining only unmapped reads. Filtered paired and unpaired reads were individually assembled using SPAdes v3.15.3 with the "--meta" option[36], followed by co-assembly of separately pooled filtered reads from rumen and feces. Contigs longer than 1,000 bp were retained for further analysis.

Filtered reads were mapped back to the contigs using BWA-MEM[35] with default parameters for metagenomic binning. Binning was performed with MaxBin2 v2.2.7[37] and MetaBAT2 v2.27[38] using default parameters, and with CONCOCT v1.1.0[39]. For CONCOCT, contigs were first split into 10 kb fragments using cut_up_fasta.py (-c 10000 -o 0 --merge_last). Coverage tables were then generated with concoct_coverage_table.py, and the program was run with default parameters. Finally, clusters for contigs > 1 kb were merged back to the original contigs using merge_cutup_clustering.py. MAG refinement was conducted with DAS Tool v1.1.4[40] using default parameters. MAG quality was assessed using CheckM v1.2.0[41], followed by collection of "medium-quality" (completeness ≥ 50% and contamination ≤ 10%) and "high-quality" MAGs (completeness ≥ 90% and contamination ≤ 5%). These MAGs were dereplicated at a 99% average nucleotide identity (ANI) threshold using dRep v3.2.2[42]. The taxonomical classification of non-redundant MAGs was conducted using GTDB-tk v2.4.0[43], with GTDB release 220. A phylogenetic tree was constructed using PhyloPhlAn v3.0.60[44] and visualized using iTOL v7.1[45].

Protein prediction for MAGs was carried out using Prodigal v2.6.3[46]. Carbohydrate-Active enZyme (CAZyme) annotation was conducted using HMMER v3.3[47] against dbCAN HMM v10 database[48] with an E-value < 1e-15 and coverage > 0.35. Polysaccharide utilization loci (PUL) were predicted using PULpy[49].

Rumen and fecal prokaryotic community

To construct reference prokaryotic genome database, we combined rumen prokaryotic genomes from the Hungate1000 project[50] and MAGs from the ruminant GITs reported in this study and previous studies[6,51–55], followed by dereplicated at a 95% ANI threshold using dRep v3.2.2[42]. The taxonomical analysis of reference genomes was conducted using GTDB-tk v2.4.0[43]. Filtered reads were mapped to the reference database using BamM (https://github.com/Ecogenomics/BamM). Reads per kilobase per million mapped reads (RPKM) values were calculated by using CoverM v.0.4.0 (https://github.com/wwood/CoverM) with '--min-covered-fraction 75 --min-read-percent-identity 95 --min-read-aligned-percent 75'. Non-metric dimensional scaling ordinations (NMDS) based on Bray-Curtis dissimilarity were performed using "vegan" package[56].

Data processing and vOTU clustering

VLP-metagenomics data were filtered and assembled following the procedure described above. Contigs longer than 5 kb were collected for further analysis. Contigs from feces and rumen were separately pooled and clustered at 95% identity and 85% coverage using CD-HIT v4.8.1[57], generating viral operational taxonomic units (vOTUs), representing species-level viral clusters. Viral identification was performed using geNomad v1.8.0[58].

The completeness of vOTUs was evaluated using CheckV v0.8.1[59]. vOTUs classified as "complete", "high-quality" (≥ 90% completeness), or "medium-quality" (50–90% completeness) were retained for downstream analyses. Protein-coding regions in vOTUs were predicted using Prodigal v2.6.3[46]. To compare fecal vOTUs, rumen vOTUs, and RefSeq database, the vOTUs were clustered using vConTACT2 v.0.9.19[60] with its ProkaryoticViralRefSeq201-Merged database into viral cluster (VC), which is approximately genus-level viral cluster. The vOTUs that were classified as "outliers" and "overlap" were identified as a unique VC. Fecal and rumen vOTUs were clustered again at 95% identity and 85% coverage to create non-redundant vOTUs (referred to as rumen and feces viruses genomes (RFVG)).

Predicted proteins of RFVG were annotated against PHROGS v4[61] using MMseqs2 v13.45111[62], and annotations with the lowest e-values (≤ 0.001) were extracted. Taxonomy prediction was conducted using geNomad[58], and lifestyle classification was performed using PhaTYP[63]. For the identification of viral AMGs, DRAM-v[64] was used. Before AMG analysis, RFVG were processed using VirSorter2[65] with the "—prep-for-dramv" option to generate the "affi-contigs.tab" files. DRAM-v was then used to annotate the sequences and distill the predicted AMGs. Genes with an auxiliary score of 1, 2, or 3 and an M flag were retained, while those with transposon regions (T flag) were excluded. Genome maps of vOTUs which have GH AMGs were visualized using CGView[66] on Proksee web server[67].

To predict host-virus interaction, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) spacers with evidence levels of 3 or 4 were extracted from reference prokaryotic genomes using CRISPRCasFinder v4.2.20[68]. The predicted CRISPR spacers were aligned against RFVG using blastn-short from BLAST + v2.9.0[69], and perfect matches (100% identity and length) were considered host-virus interactions. The host–virus interaction network was visualized using Cytoscape v3.9.1[70].

Filtered reads from VLP sequencing were mapped to RFVG using BamM v1.7.3, and RPKM values were calculated using CoverM v.0.4.0 with '--min-read-aligned-percent 95, --min-read-percent-identity 90, --min-covered-fraction 75.' RPKM values for the same VC were summed to estimate VC abundances. NMDS based on Bray-Curtis dissimilarity was performed using "vegan" R package[56].

Statical analyses

Permutational multivariate analysis of variance (PERMANOVA) was used for beta diversity with GIT site and animal as fixed effects and the number of permutations performed by default. Differences in the relative abundance of MAGs and VCs between the rumen and feces were tested using the Mann–Whitney U test. Differences were considered statistically significant at $P < 0.05$.

Results and discussion

Overview of MAGs

From whole-metagenomic sequencing, 74.0 ± 8.76 million reads per sample (mean ± SD) were generated after quality control, leading to the reconstruction of 478 MAGs with ≥ 50% completeness and ≤ 10% contamination (263 MAGs from feces: 215 MAGs from rumen). After dereplication at 99% ANI, 307 non-redundant MAGs (165 MAGs from feces: 142 MAGs from rumen) were obtained. Among them, 131 and 176 MAGs were classified as high- and medium-quality, respectively (Supplementary Table 2). The sizes of the MAGs ranged from 459 Kb to 5.0 Mb, with an average of 2.1 Mb.

After taxonomic analysis, all MAGs were classified to the family level, but 5 and 121 MAGs could not be assigned at the genus and species levels, respectively. These results confirm that the presence of putative novel genera and species in the MAGs (Fig. 1A; Supplementary Table 2). The most common MAGs belonged to Bacillota_A (66 MAGs from feces; 33 MAGs from rumen) and Bacteroidota (53 MAGs from feces; 56 MAGs from rumen), suggesting that these phyla are dominant in the GITs, consistent with the previous studies[14,54,71]. MAGs classified as Methanobacteriota (5 MAGs) were exclusively reconstructed from rumen samples (Fig. 1A, B). At the genus level, fecal MAGs were assigned to 106 genera, while rumen MAGs were classified into 64 genera. Only 24 genera were shared between fecal and rumen MAGs, indicating that distinct microbial communities are distributed in the two GITs, with many genera specific to either rumen or feces (Fig. 1C). For instance, *Alistipes* (10 MAGs), which were prevalent in the large intestine[12], and UBA737 (5 MAGs) was among the frequent genera in feces but was not reconstructed from rumen. Conversely, *Prevotella* (15 MAGs), the

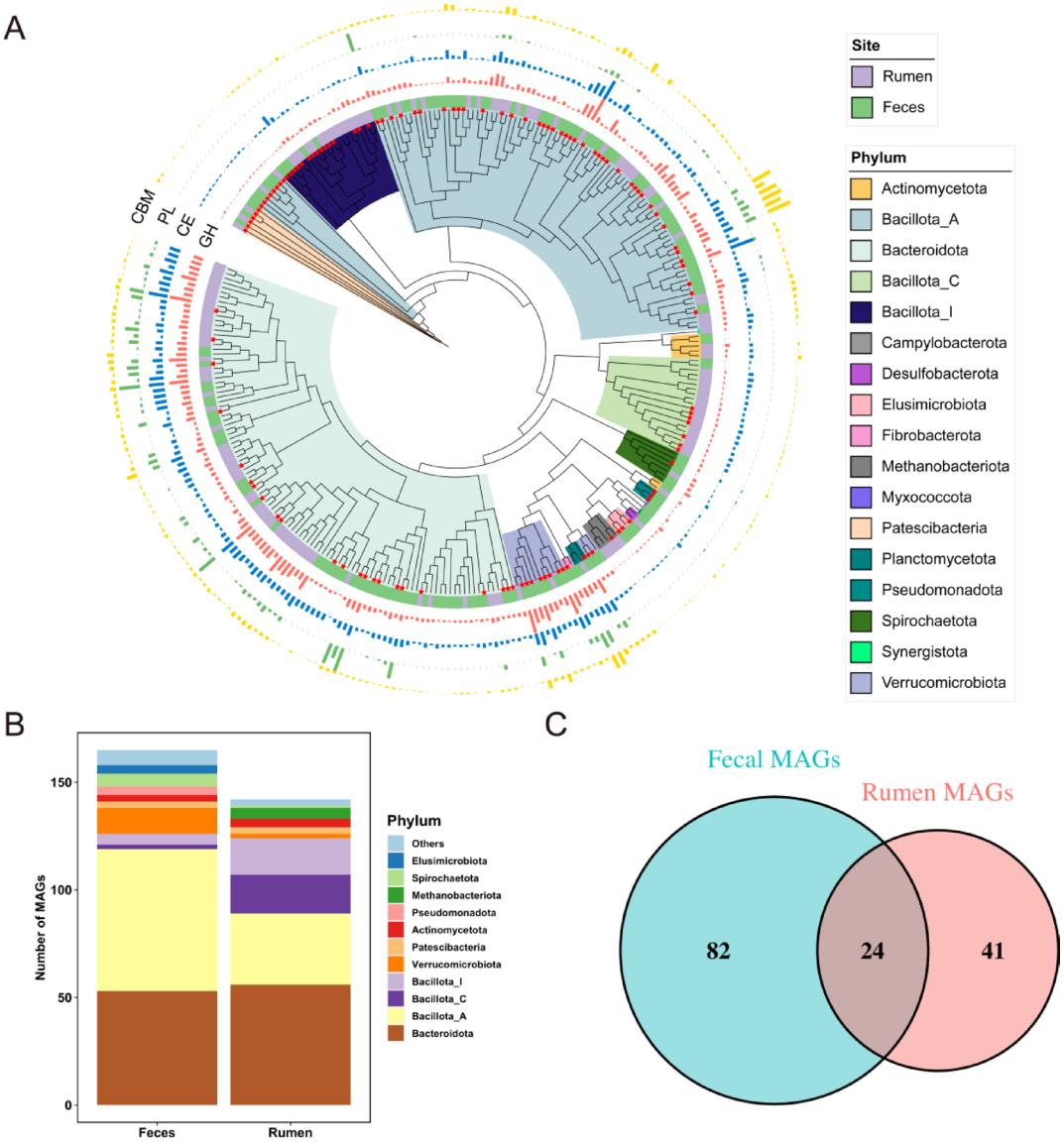

Fig. 1. Overview of reconstructed MAGs. **A** Phylogenetic tree of 307 MAGs from feces and rumen of wethers. The red star symbol represents putative novel species. **B** The number of reconstructed MAGs from feces and rumen. **C** Venn diagram showing the distribution of MAG genera from feces and rumen.

dominant genus in rumen, *Quinella* (10 MAGs), and *Sodaliphilus* (7 MAGs) were reconstructed exclusively from rumen (Supplementary Table 2). *Cryptobacteroides* MAGs, the most frequent genus in rumen, was detected in both rumen (16 MAGs) and feces (3 MAGs), though it was far less abundant in feces. These results demonstrate that although feed particles flow from the forestomach into the lower GIT, many prokaryotes do not persist during transit and are instead site-specific.

CAZymes and PULs analysis

A total of 13,223 GHs were identified in the MAGs, ranging from 0 to 217 per genome. Many fecal MAGs with high numbers of GHs belonged to Verrucomicrobiota, while rumen MAGs were affiliated with Bacteroidota including *Prevotella* and *Cryptobacteroides*, suggesting that different taxa contribute to carbohydrate digestion in different GIT sites. The Verrucomicrobiota MAGs harbored a high abundance of GHs involved in the degradation of O-linked and N-linked host glycans, such as GH20, GH29, GH109, and GH163 (Supplementary Fig. 1), consistent with a previous study[72]. Furthermore, *Alistipes* MAGs also possessed these GHs (Supplementary Fig. 1). Thus, these bacteria mainly contribute to host glycan degradation in the lower GITs. In line with this, bacterial genomes with a high abundance of GHs related to host glycan degradation have been identified in feces from sheep[55]. Additionally, these MAGs also harbored GHs involved in plant biomass degradation. For example, Verrucomicrobiota MAGs have GHs involved in hemicellulose degradation such as GH2, GH28, GH39, while some *Alistipes* MAGs possessed GH3 (β-glucosidase), GH43 (xylanase), and GH13 (α-amylase), probably indicating that the taxa also contribute to the metabolisms of cellobiose, hemicellulose and starch in the lower GIT (Supplementary Fig. 1). In ruminants, feeds are primally fermented in the rumen and plant biomass that escape degradation are fermented by microorganisms in the lower GITs. Thus, these MAGs involve in the degradation of plant biomass that remains undigested after fermentation in the upper GITs, but their primary function may appear to be the degradation of host glycans. In contrast, *Prevotella* and *Cryptobacteroides* MAGs in the rumen harbored a variety of GHs involved to carbohydrate degradation including GH5, GH9 (cellulase), GH10, GH43 (xylanase), and GH13 (α-amylase), consistent with previous studies[53,73].

For PULs analysis, 656 PULs were found in 89 Bacteroidota MAGs. The most number of PULs in MAGs were 34 (MAG244), followed by 27 (MAG54 and MAG153) (Supplementary Table 3). Those MAGs were belonging to *Cryptobacteroides* and *Prevotella*. Those genera have many PULs including GHs related to carbohydrate degradation, consistent with in the previous studies[53,74]. For example, PULs in these MAGs were including many varieties of GHs such as GH2, GH3, GH5, GH13, and GH43, related to degradation of cellulose, hemicellulose and starch. Similarly, *Alistipes* MAGs also have many PULs including these GHs, ranged from 3 to 15, (Supplementary Table 3), indicating that those contribute for plant biomass degradation in the ruminant GITs.

Overview of vOTUs

From VLP-metagenomic sequencing of rumen, 43.5 ± 12.66 and 157.1 ± 14.36 million reads per sample (mean ± SD) were generated after quality control from Method 1 and 2, respectively, while 26.5 ± 7.62 million reads per sample (mean ± SD) were generated for feces. After identification of the viral genomes and quality checks, a total of 6,701 vOTUs (1,019 fecal and 5,682 rumen vOTUs) of medium quality and above were identified. Due to the significantly low sequencing data in fecal samples compared to rumen samples, very few fecal vOTUs were collected. After clustering fecal and rumen vOTUs, 6,224 vOTUs (665 fecal and 5,559 rumen vOTUs) were produced (Supplementary Table 4), suggesting that some vOTUs were shared between the rumen and feces. The sizes of RFVG ranged from 5,441 to 466,259 bp (Supplementary Table 4). Among RFVG, 6,214 vOTUs (99.8%) were taxonomically classified as Caudoviricetes (Supplementary Table 4), indicating that most viruses represent tailed viruses, consistent with previous reports[15,17–19,23,27]. Additionally, 406 (61.1%) and 3,235 (58.2%) vOTUs in RFVG were predicted as temperate viruses, respectively (Supplementary Table 4). It is often assumed that temperate viruses are dominated in high host density environments[75,76]. Indeed, high abundance of temperate viruses were observed in the rumen[18,19] and human gut[77], which expected to high abundance of hosts. Temperate viruses integrate into their host genomes as proviruses, modulating host metabolism through auxiliary genes and horizontal gene transfer[20]. Additionally, proviruses can be induced to enter the lytic cycle by certain environmental factors, leading to host cell lysis. Thus, viruses primarily regulate host abundance and metabolism through lysogenic interactions in the ruminant GITs. However, the previous study on the ruminant GITs virome reported the inconsistent result[15,27], maybe due to the difference of the prediction procedure of virus lifestyle[20]. Further study is needed to validate the results in this study.

To compare fecal and rumen vOTUs and RefSeq viral genomes, a gene-sharing network was constructed. In the network, fecal and rumen vOTUs were clearly separated from the RefSeq, whereas both vOTUs were related to each other (Fig. 2A). The fecal and rumen vOTUs were clustered into 683 and 2,432 VCs, respectively (Fig. 2B). Among them, only 11 and 18 VCs from feces and rumen, respectively, were shared with VCs in RefSeq, suggesting that most VCs in this study had an undescribed diversity of viruses (Fig. 2B). In contrast, 255 VCs (37.3%) from feces were shared with those from rumen, indicating that viruses in the ruminant GIT are more similar to each other than to viruses from other environments (Fig. 2B).

After analyzing host-virus interactions, the hosts of 415 vOTUs (6.2% of RFVG) were predicted, with 860 links identified (Supplementary Table 5). Only 37 vOTUs in feces had predicted hosts, which is likely attributed to the reference database primarily consisting of rumen prokaryotes. Among these, 408 vOTUs interacted with specific phyla, with Bacillota_A (14 fecal and 121 rumen vOTUs) and Bacteroidota (16 fecal and 154 rumen vOTUs) being the most common hosts (Fig. 3A). These phyla are the most abundant in both the rumen and feces (Fig. 1B). At the genus level, 369 vOTUs interacted with only one genus, while 46 vOTUs were associated with multiple genera, indicating that most vOTUs are specialists, consistent with previous studies[14,19,27]. The host of vOTUs were 114 genera, suggesting a diverse range of viruses in the ruminant GIT. vOTUs interacted with *Prevotella* (65 vOTUs) were dominant, followed by *Cryptobacteroides* (27 vOTUs), and *CAG-791* (26 vOTUs)

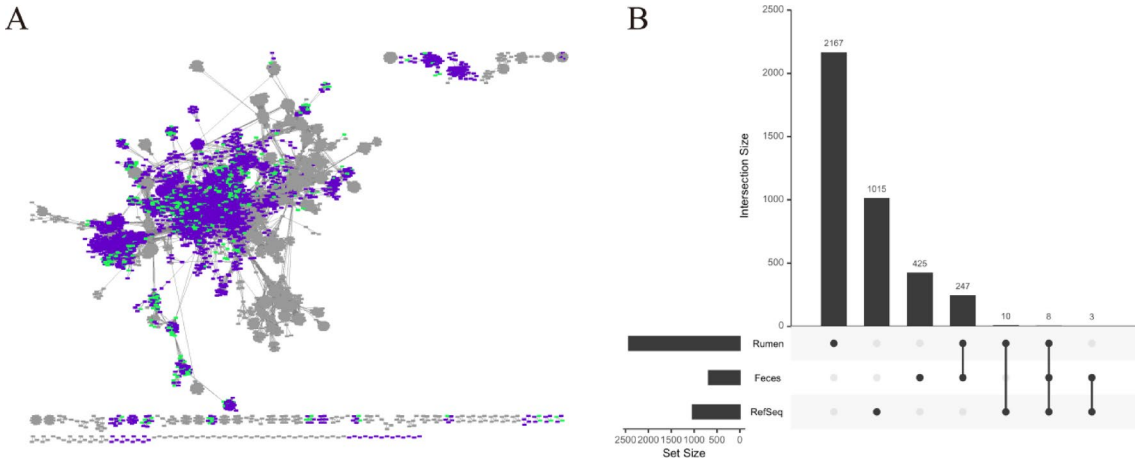

Fig. 2. Clustering of fecal and rumen vOTUs with RefSeq viral genomes based on shared genes. **A** Overview of the vConTACT2 gene-sharing network. Purple, right green, and gray nodes represent rumen, fecal, and RefSeq genomes, respectively. **B** Upset plot showing the comparison of shared viral clusters among feces, rumen, and RefSeq.

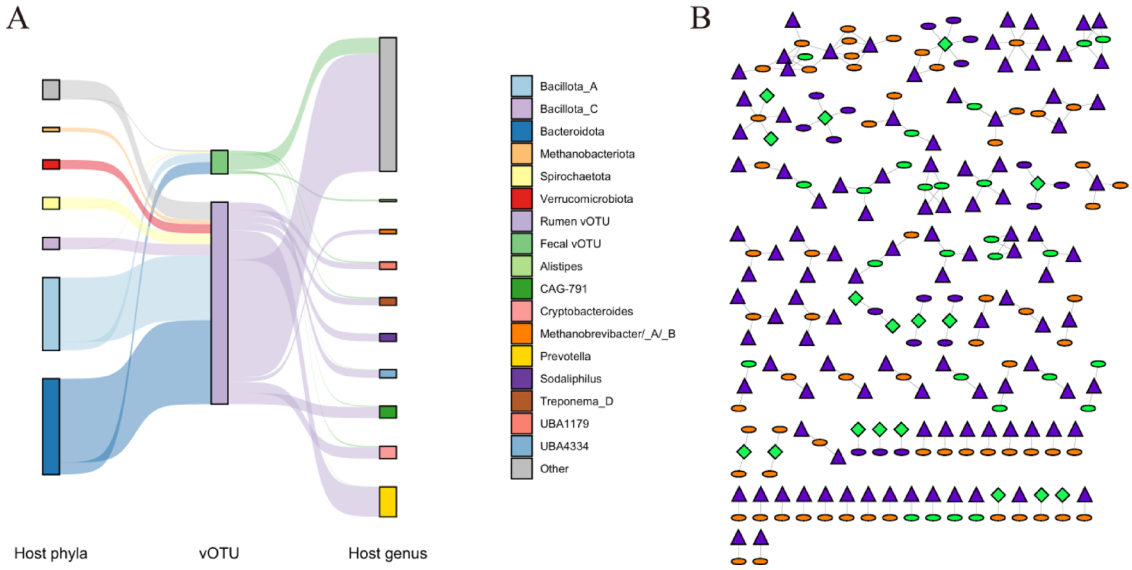

Fig. 3. Predicted host-virus interactions. **A** Sankey diagram displays the distribution of interactions between vOTUs and microbial hosts at the phyla and genus level. **B** Predicted host-virus interactions of vOTUs which linked to hosts from different sites. The shapes of triangle, diamond and circle represent rumen vOTUs, feces vOTUs and the host prokaryote, respectively. The color of right green, purple and orange represent the site of feces including rectum, rumen and other GIT sites, respectively. The network was visualized using Cytoscape[70].

(Fig. 3A). Additionally, a total of 9 rumen vOTUs were associated with methanogen such as *Methanobrevibacter*, *Methanobrevibacter_A*, and *Methanobrevibacter_B* (Fig. 3A). Most vOTUs interacting with these genera were reconstructed from rumen samples. In contrast, all vOTUs related to *Alistipes* (4 vOTUs) were from fecal samples. Next, we investigated whether the predicted hosts originated from the same GIT sites as the vOTUs. As a result, 123 vOTUs were associated with hosts from different GIT sites (Fig. 3B). Surprisingly, 21 out of 37 vOTUs identified in fecal samples were predicted to interact with hosts originating from other GIT regions, potentially indicating that these viruses flowed into the rectum from upper GIT sections.

For functional analysis, we investigated AMGs. As a result, 3,355 AMGs (425 from feces and 2,930 from rumen) were identified in 1,644 vOTUs (186 from feces and 1,456 from rumen). The most dominant AMGs were dUTP pyrophosphatase (K01520) (585 AMGs), followed by ribonucleoside-triphosphate reductase (K21636) (474 AMGs) and mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase (PF01832) (291 AMGs), which is consistent with previous rumen virome studies[17–19] (Supplementary Fig. 2A). Additionally, 10 rumen vOTUs

contained AMGs encoding GHs, including GH2, GH5, and GH16, indicating that these vOTUs may be involved in carbohydrate metabolism during infection (Supplementary Fig. 2B). For example, all GH2 AMGs were found in vOTUs belonging to VC_129, whose putative hosts are Bacteroidota, including *Prevotella*, *Enterocola*, and *Limimorpha*. Additionally, vOTU438, whose host is *Limimorpha*, had GH16 AMGs. *Prevotella* plays an essential role in carbohydrate degradation in the rumen[78]. Furthermore, MAGs belonging to *Enterocola* (MAG148 and MAG263) harbored PULs containing GH13, while *Limimorpha* (MAG159) harbored PULs containing GH3 and GH43 (Supplementary Table 3). Although the role of these genera in carbohydrate metabolism in the rumen had not been well understood, these findings suggest that they contribute to carbohydrate degradation, with the vOTUs potentially aiding this process during infection. In contrast, no GH AMGs were found in fecal vOTUs, likely due to the functional difference between the rumen and rectum in fiber digestion. The abundance of CAZyme domains was higher in the rumen than in the rectum[79]. However, the possibility that fewer vOTUs were identified from feces compared to the rumen cannot be ruled out. To fully understand the role of fecal vOTUs in carbohydrate metabolism, increasing the number of identified vOTUs is necessary.

Difference individual prokaryome and virome between fecal and rumen sample

First, we investigated whether individual viromes differed depending on the VLP purification and DNA extraction methods. The results showed that the same samples clustered together based on Bray-Curtis dissimilarity, indicating that both methods were suitable for VLP-metagenome sequencing of rumen (Supplementary Fig. 3). Method 1 is a simpler procedure compared to Method 2, suggesting that Method 1 is more suitable. Based on these results, we used VLP-metagenomic data obtained from Method 1 to investigate differences in the virome between rumen and fecal samples.

For prokaryome analysis, we used a large reference database composed of 11,356 ruminant GIT prokaryotic genomes (dereplicated at a 95% ANI threshold) for mapping. However, only 522 prokaryotes, including 147 representative MAGs reconstructed in this study, were detected in at least one sample. Additionally, the mapping rate of MAGs was 30.8 ± 8.62%, suggesting that many prokaryotes present in our samples were not included in the reference database. For example, *Fibrobacter*, which is important bacteria for cellulose degradation in the rumen[78], were not present in the rumen samples. Similarly, the mapping rate of the virome was 48.0 ± 19.02%. These results indicate that further expansion of MAG and viral genome databases is needed for a more comprehensive understanding of the prokaryome and virome.

NMDS plots of the prokaryome showed distinct clustering according to GIT site (PERMANOVA; $P = 0.001$) (Fig. 4A). The most dominant phylum in the fecal samples is Bacillota_A (46.9%), followed by Bacteroidota (35.6%) and Verrucomicrobiota (4.61%) (Supplementary Fig. 4). In contrast, in rumen samples, Bacteroidota (62.9%) is the most dominant phylum, followed by Bacillota_C (13.7%) and Bacillota_A (9.94%) (Supplementary Fig. 4). This result is broadly consistent with the number of phyla of reconstructed MAGs. The relative abundance of Bacillota_A, Verrucomicrobiota, Fibrobacterota, Pseudomonadota, and Halobacteriota was higher in feces than those in rumen, whereas Bacillota_I and Bacillota_C were higher in rumen than in feces ($P < 0.05$) (Fig. 4B). At the genus level, 22 genera were more abundant in feces than in rumen, while 7 genera were more abundant in rumen ($P < 0.05$) (Fig. 4C), indicating that the two GIT sites present significantly different prokaryotic taxa. For example, *Prevotella* and *Alistipes*, which were reconstructed exclusively from fecal or rumen samples, were more abundant in their respective GIT sites.

For virome, in line with prokaryome, NMDS plot exhibited distinct clustering between feces and rumen, whereas the prokaryome showed clearer clustering than virome (PERMANOVA; $P = 0.031$) (Fig. 5A). A total of 145 VCs showed significant differences between feces and rumen, with 6 and 139 VCs enriched in feces and rumen, respectively ($P < 0.05$) (Fig. 5B), suggesting that rumen viruses were more site-specific than fecal viruses. However, this observation may be influenced by the fact that the number of rumen vOTUs was much higher than those from feces.

Among these 145 VCs, the putative hosts of 52 VCs were predicted (Supplementary Table 6). Notably, many VCs interacting with *Prevotella* and *Sodaliphilus* were enriched in the rumen. The MAGs of these genera were reconstructed exclusively from rumen samples, and these genera were also enriched in rumen based on the prokaryome analysis (Fig. 4C). Thus, the high abundance of host genera in the rumen is accompanied by a high abundance of VCs associated with them. Additionally, the abundance of VCs related to *Methanobrevibacter* and *Methanobrevibacter_B* was also higher in the rumen. *Methanobrevibacter* is the dominant methanogen in the rumen[11], which likely explains the high abundance of VCs associated with these genera. These findings suggest that these viruses are likely rumen-specific and may contribute to carbohydrate and methane metabolism.

Next, we investigated whether the MAGs and vOTUs were present in both GIT sites to determine their site specificity, based on read mapping using CoverM. Notably, only 15 MAGs (2.87% of the detected MAGs) were shared between fecal and rumen samples (Fig. 6A), indicating that prokaryotes exhibit strong site specificity. This is consistent with the concept that distinct microbial communities with specific metabolic functions are established in each GIT compartment, where fermentation and digestion occur in a site-dependent manner[79]. In contrast, a considerably higher proportion of vOTUs (1,432; 23.0% of RFVG) were found in both fecal and rumen samples (Fig. 6B). This suggests that although many vOTUs are site-specific, some are ubiquitous across the ruminant GITs and may influence microbial ecology throughout these environments. A previous study[15] reported that over 90% of viruses (at the VC level) in the ruminant GIT were site-specific, a higher proportion than observed in our study. This discrepancy is probably attributed to the fact that the previous study used whole-metagenomic sequencing, whereas we employed VLP-metagenomics. In general, the proportion of viral reads was significantly low in whole-metagenomic sequencing, making it difficult to fully understand viral diversity in the GITs of each individual. In contrast, most reads from VLP-metagenomics were of viral origin and provided a more detailed insight compared to whole-metagenomics. Therefore, the high sensitivity of VLP-metagenomics enabled us to capture even low-abundance viruses, revealing that viral communities are more

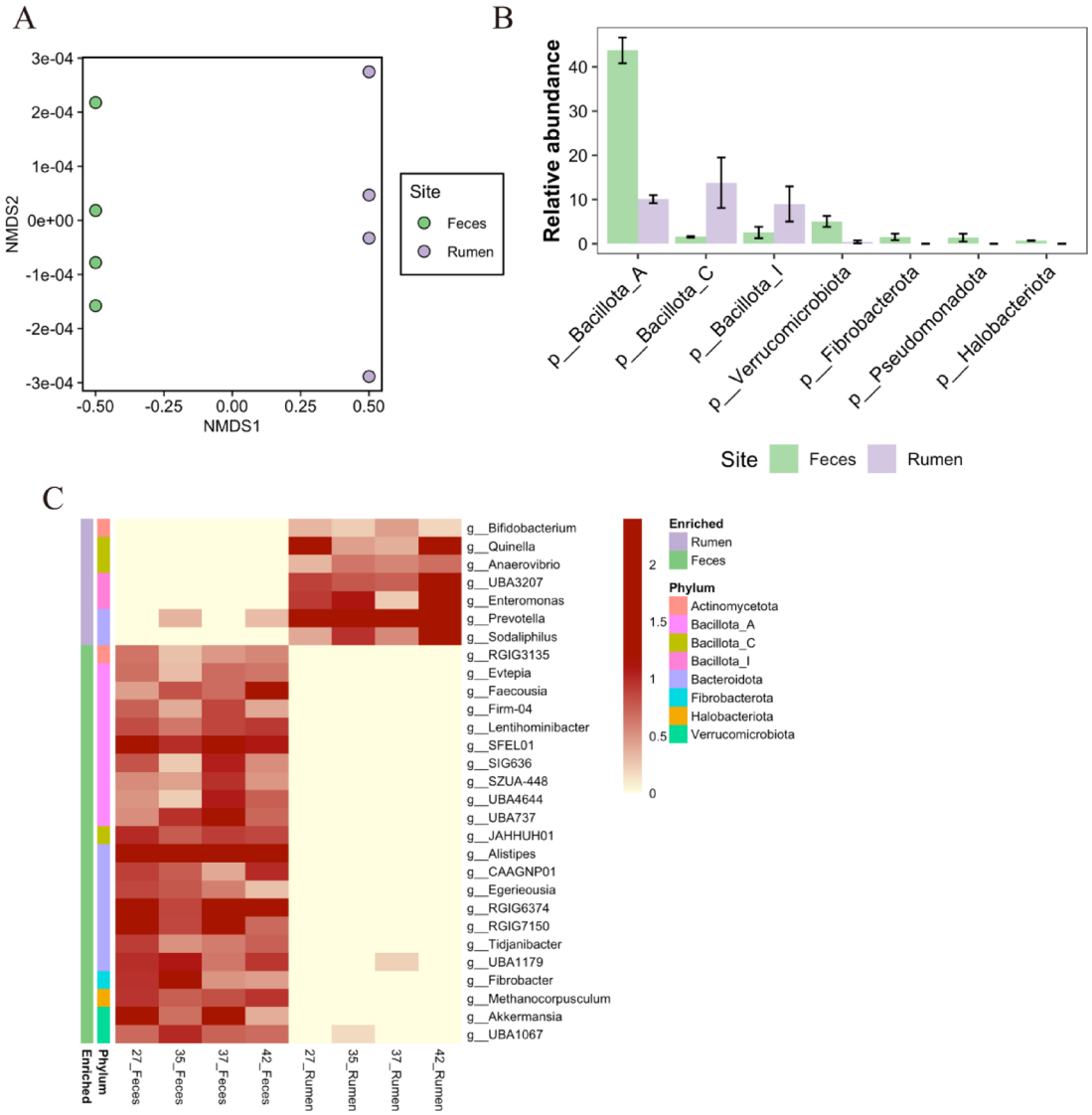

Fig. 4. Difference of prokaryotic community between fecal and rumen samples according to MAGs based analysis. **A** Non-metric multidimensional scaling ordination plot based on the Bray–Curtis dissimilarity ($P < 0.05$). **B** Significantly different phylum between two GIT sites ($P < 0.05$). **C** Heatmap depicting the significantly different genera between two GIT sites ($P < 0.05$). The values represent the logarithmically transformed RPKM values ($\log_{10}(\text{RPKM} + 1)$).

widely distributed in the GIT than previously recognized. Moreover, the majority of these vOTUs were detected in both fecal and rumen samples from the same individuals (Fig. 6B), suggesting that vOTUs in the ruminant GIT are highly individual-specific, consistent with previous studies[17–19].

Among the 1,432 vOTUs that were found both GIT sites, hosts of 47 vOTUs were predicted, with most of these viruses being specialists. We then examined whether these host genera were present in both fecal and rumen samples. Surprisingly, only 19 vOTUs host genera were detected in both GITs, while the hosts of the remaining vOTUs were exclusively found in either the rumen or feces, or were absent from both (Fig. 6C). This result demonstrates that rumen-specific viruses may pass through the lower GIT, with their DNA remaining detectable in the feces despite their inability to infect hosts in that environment.

Together, our findings suggest two possible explanations for the detection of shared vOTUs between rumen and feces. Some vOTUs may indeed be widespread across the ruminant GIT and ecologically active in multiple compartments if their hosts are present. Alternatively, certain rumen-specific viruses may simply pass through the lower GIT, with their DNA detectable in feces without infecting hosts in that environment. Our DNA-based approach cannot fully discriminate between these possibilities. To clarify whether these shared vOTUs represent ecologically relevant viral activity or passive DNA passage, future studies should also examine intermediate GIT compartments, such as the omasum, abomasum, and small intestine, in addition to rumen and feces. Furthermore, while the number of reconstructed MAGs in this study was relatively low, it cannot be ruled out

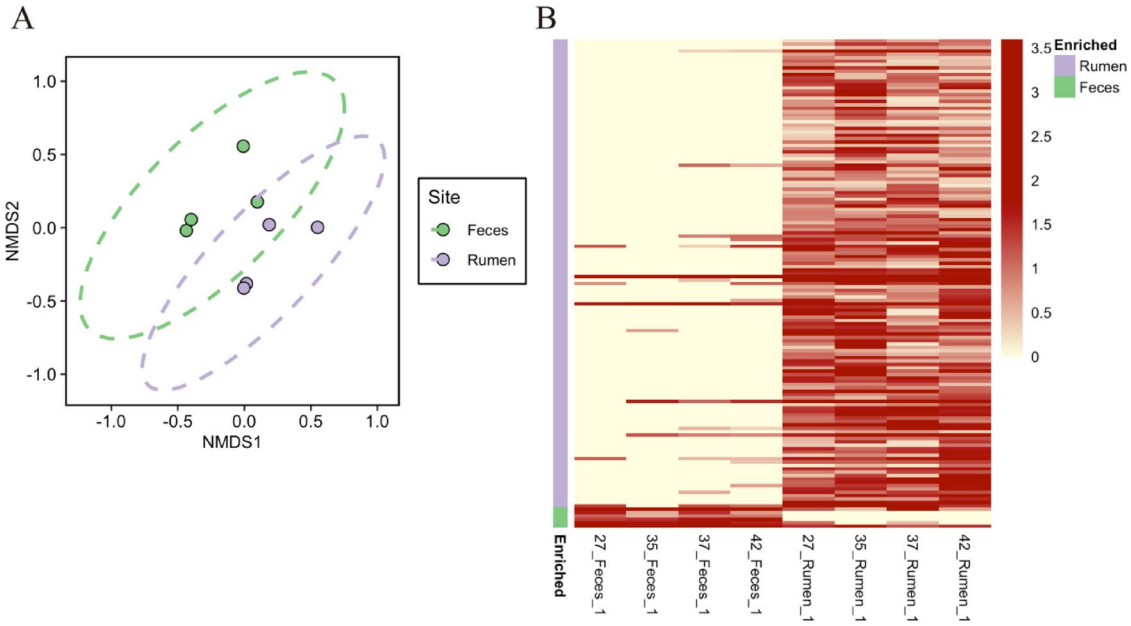

Fig. 5. **A** Difference of viral community between fecal and rumen samples. Non-metric multidimensional scaling ordination plot based on the Bray–Curtis dissimilarity ($P < 0.05$). **B** Heatmap representing the significantly different VCs between two GIT sites ($P < 0.05$). The values represent the logarithmically transformed RPKM values ($\log_{10}(\text{RPKM} + 1)$).

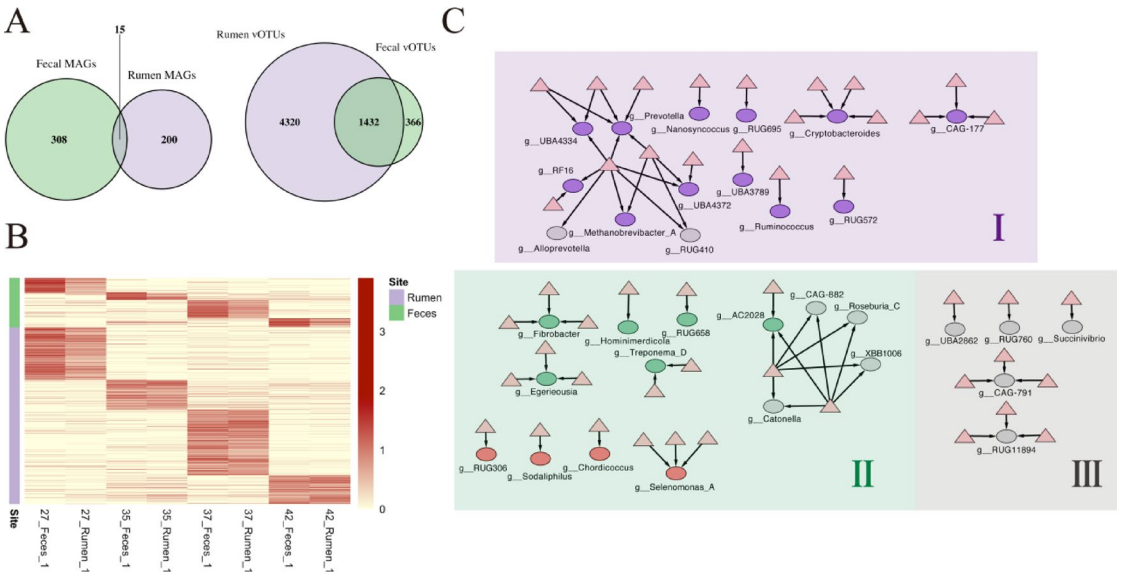

Fig. 6. Share of MAGs and vOTUs between feces and rumen. **A** Venn diagram showing the number of MAGs and vOTUs present in the rumen, feces, and those shared between both GIT sites. **B** Heatmap depicting the abundance of vOTUs found in both feces and the rumen. The values represent logarithmically transformed RPKM values ($\log_{10}(\text{RPKM} + 1)$). **C** Virus-host interaction networks based on host distribution. Cluster I, which has a purple background, represents viruses interacting with hosts that are present in both the rumen and feces. Cluster II, shown with a green background, includes viruses interacting with hosts found in either the rumen or feces but not both. Cluster III, with a gray background, represents viruses interacting with hosts that are absent from both the rumen and feces. In the network, triangles represent viruses, while circles represent microbial hosts. The color coding indicates host distribution: purple represents hosts found in both the rumen and feces, green and red indicate hosts found in the feces or rumen, respectively, and gray represents hosts that are not found in either the rumen or feces. The network was visualized using Cytoscape[70].

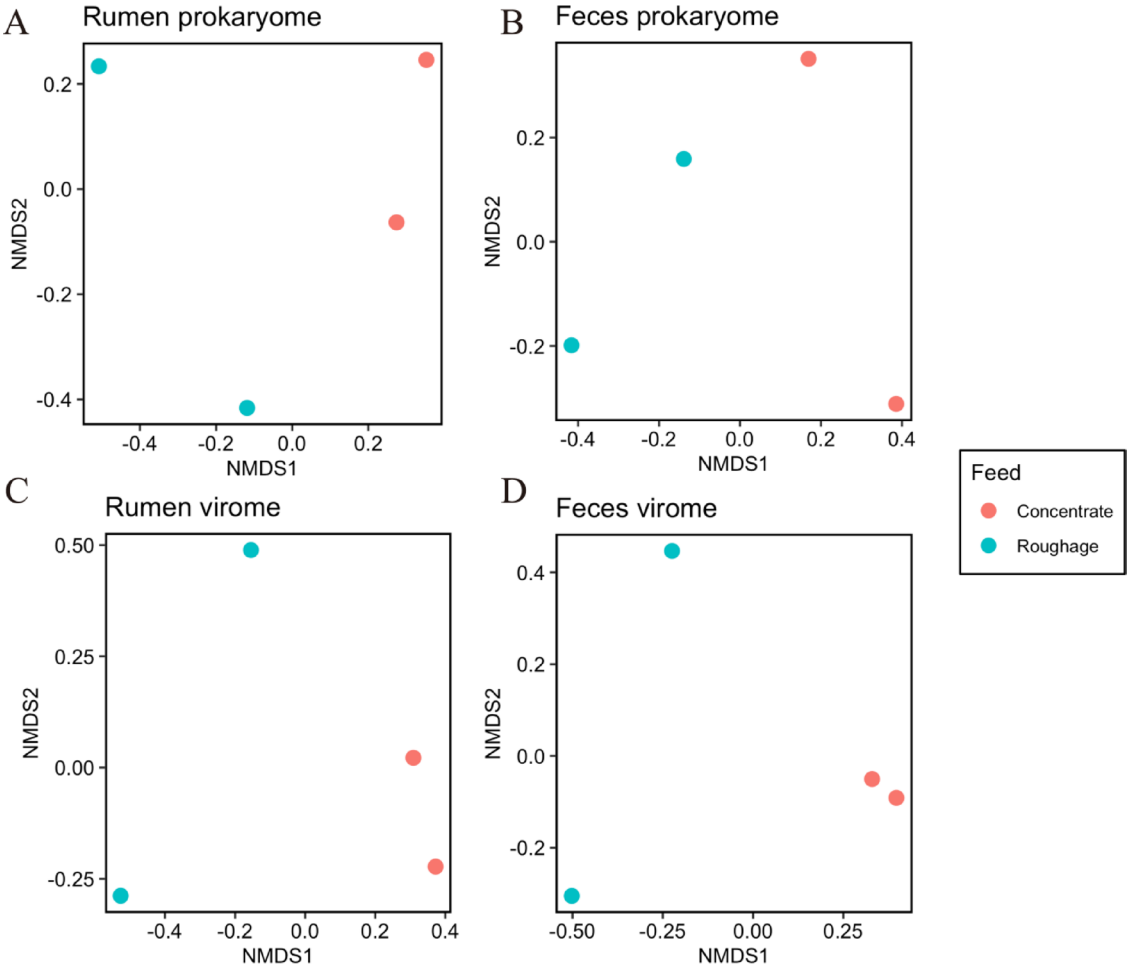

Fig. 7. Feed effects on rumen and fecal microbial diversity. NMDS plots based on Bray–Curtis dissimilarity showing **A** rumen prokaryome, **B** fecal prokaryome, **C** rumen virome, and **D** fecal virome.

that some MAGs present in the rumen or feces were not captured. Therefore, expanding the MAG dataset is crucial for improving the accuracy of host–virus identification.

Finally, we investigated the effects of diet on prokaryotic and viral communities in the rumen and feces. For beta diversity, both prokaryotic (Fig. 7A, B) and viral communities (Fig. 7C, D) tended to separate according to the concentrate-to-roughage ratio. Many previous studies have reported diet-induced changes in prokaryotic communities in the rumen[80–82] and feces[82,83]. In contrast, to our knowledge, only one study has reported diet effects on the rumen virome[16], and none have addressed the fecal virome in ruminants. Our results provide the first evidence that different concentrate-to-roughage ratios may serve as drivers of both fecal and rumen viromes, similar to the effects observed for prokaryotes. These observations suggest that diet may indirectly shape the rumen and fecal virome through its strong influence on prokaryotic host communities, which is in line with findings from other microbial ecosystems showing that substrate availability can structure both prokaryotic and viral communities[84,85]. However, given that only two animals were fed a high-concentrate diet and two animals a high-roughage diet, the small sample size ($n = 2$ per group), together with large individual variation, precludes drawing robust conclusions about diet effects. Future studies with larger cohorts are therefore required to clarify the impact of diet on the virome in the ruminant GIT, including not only different concentrate-to-roughage ratios but also other feed types and dietary compositions.

Conclusions

This study provided a comprehensive analysis of the prokaryome and virome in the rumen and feces. Whole-metagenomic sequencing identified a diverse array of prokaryotic MAGs, including several novel taxa. Notably, the prokaryotic MAGs exhibited strong site specificity, with distinct microbial compositions between feces and rumen. In parallel with the prokaryome, VLP-metagenomics revealed extensive viral diversity, with a significant proportion of novel viral vOTUs. Unlike prokaryotes, many vOTUs were shared between fecal and rumen, suggesting that rumen viruses can transit through the GITs, display broader ecological distribution, and potentially be released into the external environment. Host-virus interaction analysis indicated that these viruses infect dominant microbial hosts in the feces and rumen, including methanogens, which may play a role in shaping microbial community dynamics and host metabolism. These findings highlight the site specificity of the

prokaryome and the relatively broader distribution of viruses, underscoring the complex ecological interactions within the ruminant GITs. Further studies are needed to refine host-virus predictions, expand reference databases, and explore the functional implications of viral communities in microbial ecosystems, digestion, and animal production.

Data availability

Raw sequence data were deposited in DDBJ (accession number: PRJDB20338). non-redundant MAGs and vOTUs are available in figshare (https://doi.org/10.6084/m9.figshare.28714169.v1).

Received: 14 July 2025; Accepted: 3 November 2025
Published online: 05 December 2025

References

1. Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* **70**, 567–590 (1990).
2. Jami, E., White, B. A. & Mizrahi, I. Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *PLoS One* **9**, e85423 (2014).
3. Xue, M-Y., Sun, H-Z., Wu, X-H., Liu, J-X. & Guan, L. L. Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. *Microbiome* **8**, 64 (2020).
4. Kim, M., Park, T., Jeong, J. Y., Baek, Y. & Lee, H-J. Association between rumen microbiota and marbling score in Korean native beef cattle. *Animals* **10**, 712 (2020).
5. Krause, T. R. et al. The relationship between the rumen microbiome and carcass merit in Angus steers. *J. Anim. Sci.* **98**, skaa287 (2020).
6. Sato, Y., Sato, R., Fukui, E. & Yoshizawa, F. Impact of rumen microbiome on cattle carcass traits. *Sci. Rep.* **14**, 6064 (2024).
7. Myer, P. R., Smith, T. P. L., Wells, J. E., Kuehn, L. A. & Freetly, H. C. Rumen microbiome from steers differing in feed efficiency. *PLoS One* **10**, e0129174 (2015).
8. Shabat, S. K. et al. Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J.* **10**, 2958–2972 (2016).
9. Shi, W. et al. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res.* **24**, 1517–1525 (2014).
10. Tapio, I., Snelling, T. J., Strozzi, F. & Wallace, R. J. The ruminal microbiome associated with methane emissions from ruminant livestock. *J. Anim. Sci. Biotechnol.* **8**, 7 (2017).
11. Janssen, P. H. & Kirs, M. Structure of the archaeal community of the rumen. *Appl. Environ. Microbiol.* **74**, 3619–3625 (2008).
12. Lin, L., Lai, Z., Zhang, J., Zhu, W. & Mao, S. The gastrointestinal microbiome in dairy cattle is constrained by the deterministic driver of the region and the modified effect of diet. *Microbiome* **11**, 10 (2023).
13. Myer, P. R., Freetly, H. C., Wells, J. E., Smith, T. P. L. & Kuehn, L. A. Analysis of the gut bacterial communities in beef cattle and their association with feed intake, growth, and efficiency. *J. Anim. Sci.* **95**, 3215–3224 (2017).
14. Cao, Y. et al. The multi-kingdom microbiome of the goat gastrointestinal tract. *Microbiome* **11**, 219 (2023).
15. Wu, Y. et al. A compendium of ruminant gastrointestinal phage genomes revealed a higher proportion of lytic phages than in any other environments. *Microbiome* **12**, 69 (2024).
16. 16son CL, Sullivan, M. B. & Fernando, S. C. Dietary energy drives the dynamic response of bovine rumen viral communities. *Microbiome* **5**, 155 (2017).
17. Yan, M. et al. Interrogating the viral dark matter of the rumen ecosystem with a global Virome database. *Nat. Commun.* **14**, 5254 (2023).
18. Sato, Y. et al. A rumen virosphere with implications of contribution to fermentation and methane production, and endemism in cattle breeds and individuals. *Appl. Environ. Microbiol.* **90**, e01581–e01523 (2024).
19. Sato, Y. Rumen DNA Virome in beef cattle reveals an unexplored diverse community with potential links to carcass traits. *ISME Commun.* **5**, ycaf021 (2025).
20. Yu, Z., Somasundaram, S. & Yan, M. Rumen protozoa and viruses: New insights into their diversity and potential roles through omics lenses—a review. *J. Dairy. Sci.* **108**, 7530-7544 (2025).
21. Paynter, M. J. B., Ewert, D. L. & Chalupa, W. Some morphological types of bacteriophages in bovine rumen contents. *Appl. Microbiol.* **18**, 942–943 (1969).
22. Klieve, A. V. & Bauchop, T. Morphological diversity of ruminal bacteriophages from sheep and cattle. *Appl. Environ. Microbiol.* **54**, 1637–1641 (1988).
23. Berg Miller, M. E. et al. Phage–bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environ. Microbiol.* **14**, 207–227 (2012).
24. Solden, L. M. et al. Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. *Nat. Microbiol.* **3**, 1274–1284 (2018).
25. Zhong, Z-P. et al. Viral potential to modulate microbial methane metabolism varies by habitat. *Nat. Commun.* **15**, 1857 (2024).
26. Yan, M. & Yu, Z. Viruses contribute to microbial diversification in the rumen ecosystem and are associated with certain animal production traits. *Microbiome* **12**, 82 (2024).
27. Liu, X., Tang, Y., Chen, H., Liu, J-X. & Sun, H-Z. Rumen DNA virome and its relationship with feed efficiency in dairy cows. *Microbiome* **13**, 14 (2025).
28. Gómez-Doñate, M. et al. Isolation of bacteriophage host strains of *Bacteroides* species suitable for tracking sources of animal faecal pollution in water. *Environ. Microbiol.* **13**, 1622–1631 (2011).
29. Gregory, A. C. et al. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell. Host Microbe.* **28**, 724–740 (2020).
30. Sato, Y. et al. Taxonomic and functional characterization of the rumen microbiome of Japanese black cattle revealed by 16S rRNA gene amplicon and metagenome shotgun sequencing. *FEMS Microbiol. Ecol.* **97**, fiab152 (2021).
31. Nishijima, S. et al. Extensive gut Virome variation and its associations with host and environmental factors in a population-level cohort. *Nat. Commun.* **13**, 5252 (2022).
32. Hurwitz, B. L., Deng, L., Poulos, B. T. & Sullivan, M. B. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ. Microbiol.* **15**, 1428–1440 (2013).
33. Tillett, D. & Neilan, B. A. Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. *J. Phycol.* **36**, 251–258 (2000).
34. Bolger, A. M., Lohse, M., & Usadel, B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
35. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 13033997 (2013).

36. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
37. Wu, Y-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated Binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
38. Kang, D. D. et al. MetaBAT 2: an adaptive Binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
39. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods*. **11**, 1144–1146 (2014).
40. Sieber, C. M. K. et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).
41. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
42. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
43. Chaumeil, P-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
44. Asnicar, F. et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* **11**, 2500 (2020).
45. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
46. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
47. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121–e121 (2013).
48. Zhang, H. et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
49. Stewart, R. D., Auffret, M. D., Roehe, R. & Watson, M. Open prediction of polysaccharide utilisation loci (PUL) in 5414 public Bacteroidetes genomes using PULpy. *bioRxiv* https://doi.org/10.1101/421024(2018).
50. Seshadri, R. et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 collection. *Nat. Biotechnol.* **36**, 359–367 (2018).
51. Anderson, C. L. & Fernando, S. C. Insights into rumen microbial biosynthetic gene cluster diversity through genome-resolved metagenomics. *Commun. Biol.* **4**, 818 (2021).
52. Stewart, R. D. et al. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* **37**, 953–961 (2019).
53. Sato, Y. et al. Identification of 146 metagenome-assembled genomes from the rumen microbiome of cattle in Japan. *Microbes Environ.* **37**, ME22039 (2022).
54. Xie, F. et al. An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome* **9**, 137 (2021).
55. Zhang, K. et al. Compendium of 5810 genomes of sheep and goat gut microbiomes provides new insights into the glycan and mucin utilization. *Microbiome* **12**, 104 (2024).
56. Oksanen, J. et al. The vegan package. *Community Ecol. Package*. **10**, 719 (2007).
57. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
58. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2024).
59. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
60. Bin Jang, H. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
61. Terzian, P. et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom. Bioinform.* **3**, lqab067 (2021).
62. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
63. Shang, J., Tang, X. & Sun, Y. PhaTYP: predicting the lifestyle for bacteriophages using BERT. *Brief. Bioinform*. **24**, bbac487 (2023).
64. Shaffer, M. et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* **48**, 8883–8900 (2020).
65. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
66. Stothard, P., Grant, J. R. & Van Domselaar, G. Visualizing and comparing circular genomes using the CGView family of tools. *Brief. Bioinform.* **20**, 1576–1582 (2019).
67. Grant, J. R. et al. Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res.* **51**, W484–W492 (2023).
68. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
69. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
70. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
71. Conteville, L. C. et al. Recovery of metagenome-assembled genomes from the rumen and fecal microbiomes of *Bos indicus* beef cattle. *Sci. Data*. **11**, 1385 (2024).
72. Zhou, S. et al. Characterization of metagenome-assembled genomes and carbohydrate-degrading genes in the gut microbiota of Tibetan pig. *Front. Microbiol.* **11**, 595066 (2020).
73. Lin, L. et al. Lignocellulolytic microbiomes orchestrating degradation cascades in the rumen of dairy cattle and their diet-influenced key degradation phases. *Anim. Adv.* **1**, e002 (2024).
74. Kou, X. et al. Insights into the donkey hindgut microbiome using metagenome-assembled genomes. *Animals* **14**, 3625 (2024).
75. Silveira, C. B. & Rohwer, F. L. Piggyback-the-Winner in host-associated microbial communities. *npj Biofilms Microbiomes*. **2**, 16010 (2016).
76. Knowles, B. et al. Lytic to temperate switching of viral communities. *Nature* **531**, 466–470 (2016).
77. Reyes, A. et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
78. Stewart, C. S., Flint, H. J. & Bryant, M. P. The rumen microbial ecosystem. In: (eds Hobson, P. N. & Stewart, C. S.) The Rumen Bacteria, 10–72 (Springer 1997).
79. Tong, F. et al. The microbiome of the buffalo digestive tract. *Nat. Commun.* **13**, 823 (2022).
80. Zhang, J. et al. Effect of dietary forage to concentrate ratios on dynamic profile changes and interactions of ruminal microbiota and metabolites in Holstein heifers. *Front. Microbiol.* **8**, 2206 (2017).
81. Chen, H., Wang, C., Huasai, S. & Chen, A. Effects of dietary forage to concentrate ratio on nutrient digestibility, ruminal fermentation and rumen bacterial composition in Angus cows. *Sci. Rep.* **11**, 17023 (2021).

82. Wang, S. et al. Effect of dietary concentrate-to-forage ratios during the cold season on slaughter performance, meat quality, rumen fermentation and gut microbiota of Tibetan sheep. *Animals* **14**, 3305 (2024).
83. Zhang, J. et al. Effect of limit-fed diets with different forage to concentrate ratios on fecal bacterial and archaeal community composition in Holstein heifers. *Front. Microbiol.* **9**, 976 (2018).
84. Schulfer, A. et al. Fecal viral community responses to high-fat diet in mice. *mSphere* **5**, e00833–e00819 (2020).
85. Howard, A., Carroll-Portillo, A., Alcock, J. & Lin, H. C. Dietary effects on the gut phageome. *Int. J. Mol. Sci.* **25**, 8690 (2024).

Acknowledgements
Supercomputing resources were provided by the Human Genome Center at the Institute of Medical Science, University of Tokyo, Japan.

Author contributions
Yoshiaki Sato: Conceptualization, investigation, formal analysis, supervision, funding acquisition, visualization, and writing (original draft). Hajime Kumagai: Resources, supervision and writing (review and editing). Hiroyuki Hirooka: Writing (review and editing). Takashi Yoshida: Resources and writing (review and editing).

Funding
This study was partly supported by grants from the Kieikai Research Foundation to YS.

Declarations

Competing interests
The authors declare no competing interests.

Ethics declarations
The experiment was approved by the Kyoto University Animal Ethics Committee (Permit Number: R3-36). All experimental procedures were performed in accordance with the university's guidelines and regulations approved by Kyoto University Animal Ethics Committee. This study is reported in accordance with ARRIVE guidelines.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-27351-9.

Correspondence and requests for materials should be addressed to Y.S.