



OPEN Exploration of mobile genetic elements in the ruminal microbiome of Nellore cattle

Camila A. Faleiros¹, Alanne T. Nunes¹, Osiel S. Gonçalves², Pâmela A. Alexandre³, Mirele D. Poleti¹, Elisângela C. Mattos¹, Flavio Perna-Junior⁴, Paulo H. Mazza Rodrigues⁴ & Heidge Fukumasu^{1✉}

Metagenomics has made it feasible to elucidate the intricacies of the ruminal microbiome and its role in the differentiation of animal production phenotypes of significance. The search for mobile genetic elements (MGEs) has taken on great importance, as they play a critical role in the transfer of genetic material between organisms. Furthermore, these elements serve a dual purpose by controlling populations through lytic bacteriophages, thereby maintaining ecological equilibrium and driving the evolutionary progress of host microorganisms. In this study, we aimed to identify the association between ruminal bacteria and their MGEs in Nellore cattle using physical chromosomal links through the Hi-C method. Shotgun metagenomic sequencing and the proximity ligation method ProxiMeta were used to analyze DNA, getting 1,713,111,307 bp, which gave rise to 107 metagenome-assembled genomes from rumen samples of four Nellore cows maintained on pasture. Taxonomic analysis revealed that most of the bacterial genomes belonged to the families *Lachnospiraceae*, *Bacteroidaceae*, *Ruminococcaceae*, *Saccharofermentanaceae*, and *Treponemataceae* and mostly encoded pathways for central carbon and other carbohydrate metabolisms. A total of 31 associations between host bacteria and MGE were identified, including 17 links to viruses and 14 links to plasmids. Additionally, we found 12 antibiotic resistance genes. To our knowledge, this is the first study in Brazilian cattle that connect MGEs with their microbial hosts. It identifies MGEs present in the rumen of pasture-raised Nellore cattle, offering insights that could advance biotechnology for food digestion and improve ruminant performance in production systems.

Advancements in metagenomics in rumen environments have enabled us to identify and classify the taxonomy and functional capabilities of bacteria that specialize in digesting plant material, which is the primary food source for ruminants. Given the impact of microorganisms on host cattle performance, microbiota modulation is crucial for developing a sustainable and productive livestock system¹. For instance, feed efficiency is a trait with implications for economic and environmental sustainability² and is strongly associated with host animal microbiota profiles^{3–5}. Ruminal methane production, which results from microbial fermentation in the rumen, is another example of trait that can be explained by genetic variation and the composition of the host animal microbiota^{6–8}. Approximately 65% of rumen prokaryotic species have been classified and described in taurine breeds⁹; however, less information is available on Indian cattle, which are the majority in the world, especially in Brazil.

Understanding the role of specific phylotypes in the rumen during fermentation and identifying new enzymes through metagenomic analysis can improve biotechnology for food digestion and ruminant performance in production systems¹⁰. Important and few characterized components of the rumen are the Mobile Genetic Elements (MGEs), they possess great potential for manipulation and biotechnological applications owing to their ability to confer benefits to their hosts and participate in crucial functional modifications of their hosts, carrying auxiliary metabolic genes such as glycosidic hydrolases (GH), which contribute to the breakdown of complex carbohydrates¹¹.

¹Department of Veterinary Medicine, School of Animal Science and Food Engineering (FZEA), University of São Paulo, Pirassununga, SP 13635-900, Brazil. ²Department of Microbiology, Institute of Biotechnology Applied to Agriculture (BIOAGRO), Federal University of Viçosa, Viçosa, MG 36570-000, Brazil. ³Commonwealth Scientific and Industrial Research Organization (CSIRO), Agriculture and Food, Brisbane, QLD, Australia. ⁴Department of Animal Nutrition and Production, School of Veterinary Medicine and Animal Science, University of São Paulo (FMVZ-USP), Pirassununga, São Paulo 13635-900, Brazil. ✉email: fukumasu@usp.br

In addition, viruses carry genes that contribute positively their hosts in the metabolism of nutrients and stages of the biogeochemical cycles carried out by the hosts^{12,13}. They are classified as mobile DNA elements that move within and between bacterial cells through agents, such as plasmids, bacteriophages, and transposons, resulting in the introduction of accessory genes that contribute to the evolution of host cells^{14,15}. Solden et al.¹⁶ reported that viruses play a crucial role in controlling the ecosystem functions in the rumen, they have metabolic interdependence with their microbial hosts, carry genes that can redirect carbon metabolism, and infect dominant microorganisms that degrade carbohydrates.

Early studies on bacteriophages in the rumen revealed that they can eliminate specific bacteria¹⁷. Recent research has indicated the potential use of lytic phages to regulate microorganisms in the rumen responsible for an increase in acidosis and methane production^{18,19}. This suggests that the components in the rumen are interconnected with viruses that infect ruminal bacteria, the microbiome, and the host animal interacting and affecting each other. This can change the ruminal function and the nutritional capacity of the host animal²⁰.

Recent studies have employed the Hi-C approach, which captures intimate interactions between bacterial chromosomes within living cells, to investigate MGEs in microbial communities. This method has been used in conjunction with metagenomic sequencing and has proven effective in assembling genomes and uncovering new relationships between hosts and MGEs^{21–24}. The objective of this study was to determine the existence of MGEs in the rumen content of Nellore cattle in Brazil, using the Hi-C method to establish a link between these elements and their microbial hosts.

Results

Metagenome of bovine rumen content samples

In this study, the metagenomic sequencing was performed with DNA extracted from a pool of rumen content samples of four Nellore cows, applying ProxiMeta, a proximity linkage method based on Hi-C. This method used contigs derived from shotgun sequencing and reads from a library Hi-C, and generated an assembly length of 1,713,111,307 bp, with 834,164 contigs total and 16,035 contigs in clusters allowing the assembly of 107 genomes (Fig. 1A). The most completely assembled genome was bin_4 (99.30%), which corresponded to *Treponema D. sp016288205*. Other genomes classified as complete, with completeness greater than 95%, were *UBA2868 sp003535955* (bin_1), *Treponema D. sp017421145* (bin_9), *Treponema D. sp902783155* (bin_6) and *UBA2868* (bin_12). The clusters classified as excellent (> 90%) were dominated by species from the *Lachnospiraceae* family, namely *UBA3766 sp902803015* (bin_22), *UBA1712* (bin_14), *UBA2868* (bin_3), *UBA2868 sp003535955* (bin_17), *NK4A144 sp902783395* (bin_7), *Acetatifactor sp900066565* (bin_21), *UBA1711 sp900317125* (bin_11), and *UBA1711 sp001543385* (bin_18). Clusters belonging to the families *Saccharofermentanaceae* with *Saccharofermentans* (bin_19), and *Bacteroidaceae* with *Prevotella sp000702825* (bin_5) also had an excellent completeness. The clusters considered good are presented in the Supplementary Table 1, and the 107 clusters subjected to taxonomic classification are represented in Fig. 1B.

Functional and carbohydrate-active enzymes (CAZymes) characterization of bins

Before analyzing the MGEs, we conducted a functional analysis to understand the functional characteristics encoded within metagenome-assembled genomes (MAGs) from the ruminal microbiome of Nellore cattle. For this, MAGs with completeness > 50% and those associated with MGEs were used, totaling 43 genomes (Supplementary Tables 8 and 9). This step provided essential insights into the interplay between microbial communities and the functional potential associated with MGEs in the rumen ecosystem. Our study evaluated the involvement of rumen microorganisms in these biogeochemical pathways, and we found that the majority were associated with the stages of the carbon, sulfur, nitrogen and iron cycles (Fig. 2). Regarding the carbon cycle, most of the genomes identified demonstrated involvement in the acetogenesis process and carbon degradation complex. More specifically, we identified two genomes belonging to *Lachnospiraceae* and *Bacteroidales* encoding

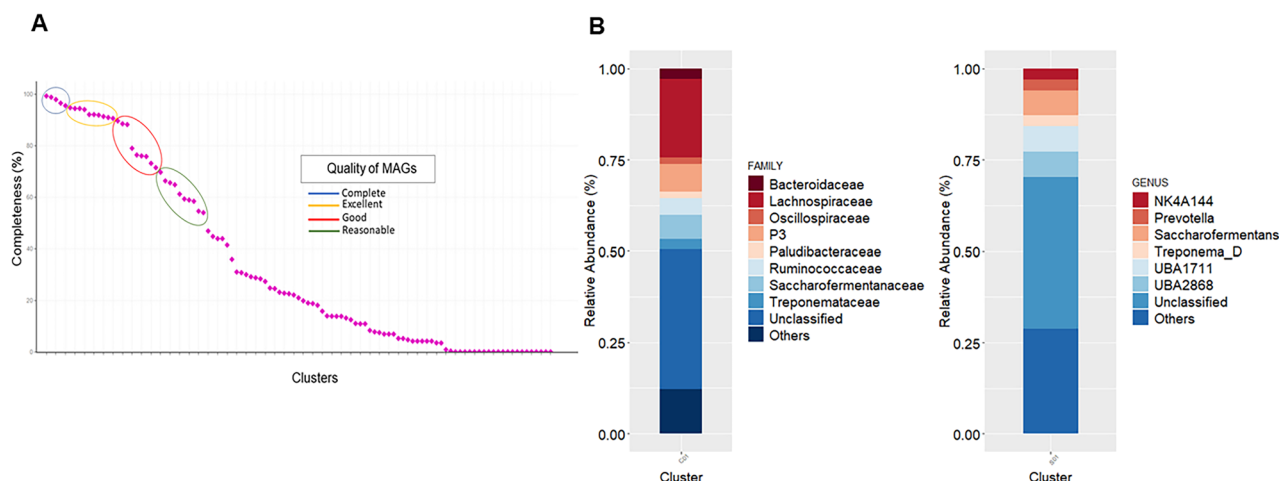


Figure 1. Rumen metagenome of Nellore cattle. (A) Completeness of metagenome-assembled genomes (MAGs), (B) Taxonomic classification of bins: graph on the left genus and right graph family.

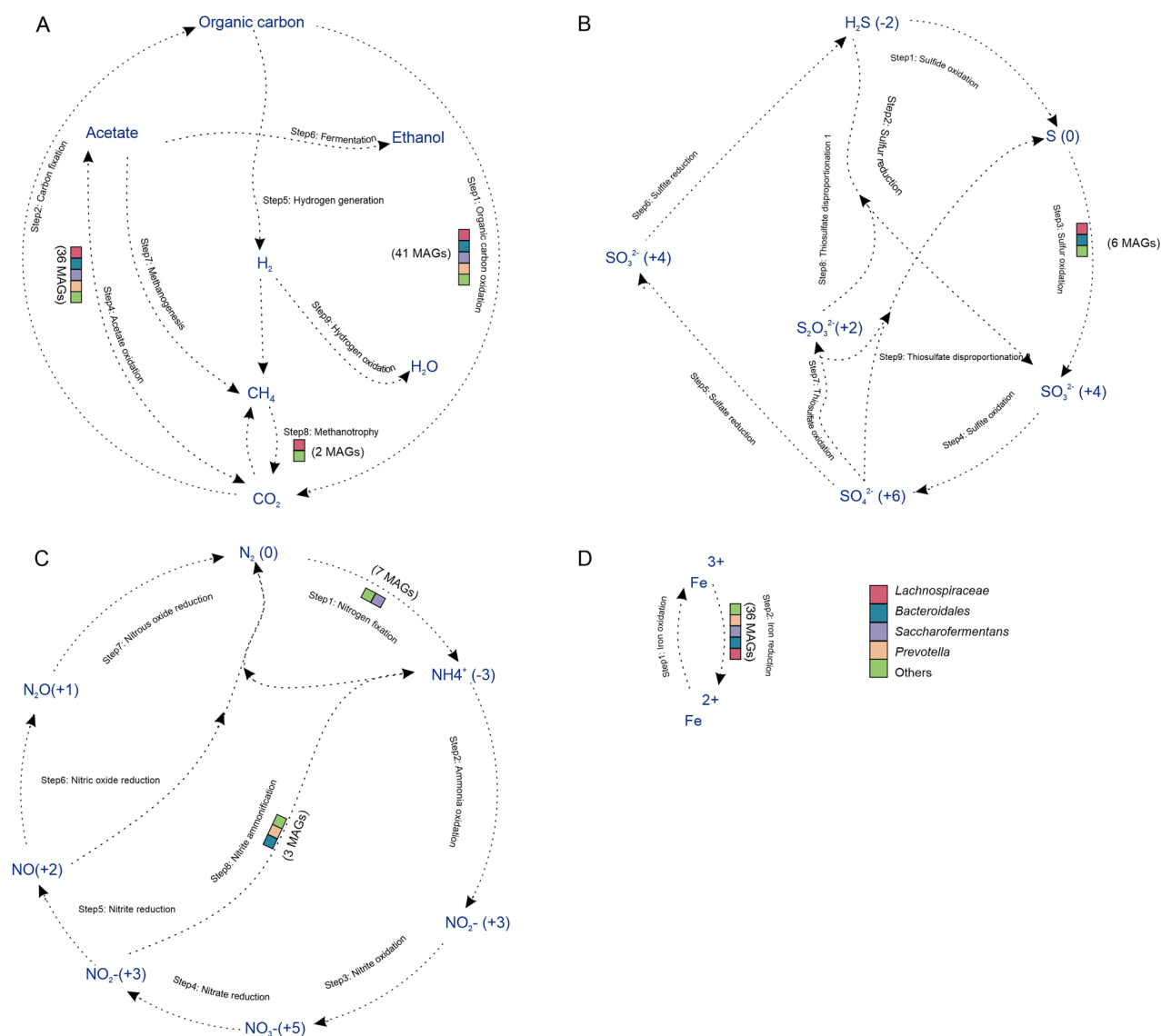


Figure 2. Potential biogeochemical cycling processes by rumen bacteria. Bins are color-coded in each step within the (A) Carbon, (B) sulfur, (C) nitrogen, and (D) iron cycles. Each arrow in the figure represents a single transformation step within a cycle.

the enzyme methane monooxygenase (mmoBD), which act in methane oxidation, and members of these clades also participated in the oxidation step of the sulfur cycle. In the nitrogen cycle, phylotypes belonging to *Saccharofermentans* and *Treponema* acted in the nitrogen fixation stage and members of the genus *Prevotella* encoded genes for the reduction of nitrite into ammonia. All taxonomic classes presented members active in the iron oxidation stage, which is fundamental for most microorganisms, due to its redox properties, being used as an electron transporter by bacteria^{25,26}.

The most frequent processes included steps in the central carbon metabolism pathway, the metabolism of other carbohydrates, and purine metabolism. Among the bins examined, 69% contained genes that encoded the phosphate acetyltransferase-acetate kinase pathway, a reversible process where acetyl-CoA can be converted into acetate or acetate into acetyl-CoA. Additionally, the aromatic amino acid metabolism pathway with tryptophan biosynthesis was present in 64% of the bins (Fig. 3A).

The metagenomic approach allows for the evaluation of the potential of rumen microbial populations to break down plant lignocellulolytic materials, by annotating the genes associated with this process, known as CAZymes¹⁰. CAZymes are classified into several categories based on their degradation capabilities, including glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), and others²⁷. In this study, the bins identified were assessed for their ability to encode CAZymes, and the presence of GHs and PLs was confirmed (Supplementary Table 3). We found that GH2, GH3, GH5, GH13, and GH43 were the most abundant enzymes mapped in the bins (Fig. 3B).

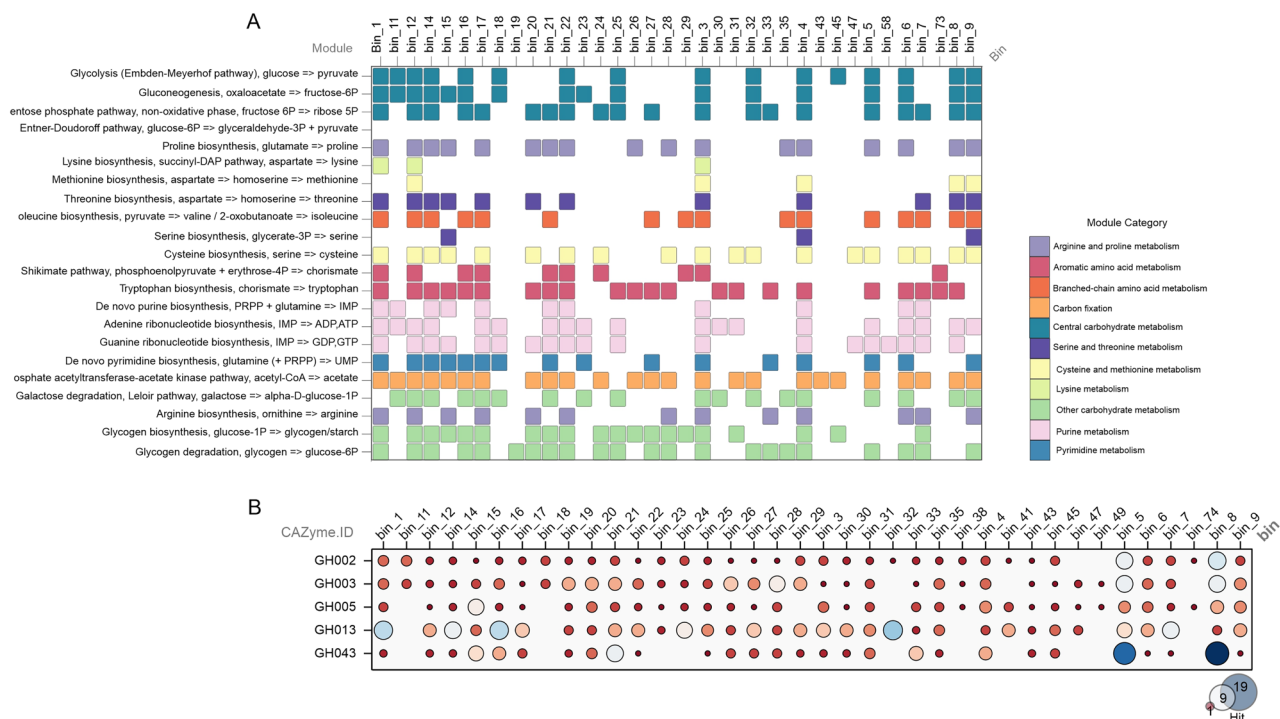


Figure 3. Central processes of metabolism and CAZymes present in the rumen microbiome of Nellore cattle. (A) Characterization of the metabolic potential of bins found in the rumen, (B) Identification and abundance of enzymes encoded by rumen microorganisms. The size of each circle corresponds to the number of hits for the respective gene.

Assignment of mobile genetic elements to their hosts

Clustering with Hi-C allows not only the construction of MAGs but also the sequence grouping of mobile genetic elements with their core genome²⁸. The sequences were obtained with sequencing and the use of the proximity linkage method to assign viruses and plasmids to their microbial hosts, the data obtained is summarized in Fig. 4, more details about the data obtained are in the supplementary material (Tables 4, 6, 10, 11 and 12).

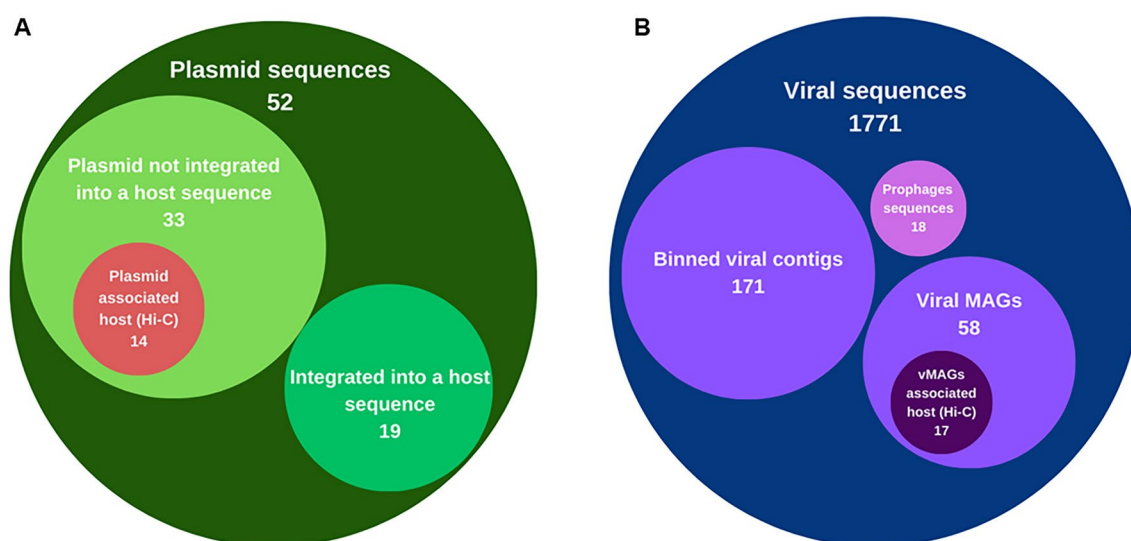


Figure 4. Mobile genetic elements of bovine rumen. (A) Each circle represents number of sequences identified as plasmids, integrated and non-host-integrated plasmids and plasmids associated with their hosts by the Hi-C method. (B) Each circle represents the number of identified viral sequences, viral sequences integrated into the host genome (Prophages), binned viral contigs, metagenome-assembled viral genomes (MAGs) and vMAGs associated with hosts by proximity linkage (Hi-C).

Plasmids and the identification of microbial resistance genes

The reads derived from shotgun sequencing were used to evaluate the presence of plasmid sequences. Thirty-three DNA sequences were identified as plasmids, which were not integrated into the host genome. Among these, 14 had at least one connection with the prokaryotic host genome through Hi-C reads. Furthermore, two putative plasmids contigs k141_1760976 and k141_5289202 were shared between multiple host clusters (bin_3, bin_5, bin_20, bin_80) and (bin_1; bin_3; bin_11; bin_30; bin_88) (Supplementary Table 4). The bins shared by contig k141_1760976 belong to different genus, including *UBA2868*, *Prevotella* and *Saccharofermentans*. The plasmid k141_5289202 shared by three individuals belonging to *Lachnospiraceae* and one belonging to *Bacteroidales*.

Additionally, we identified 19 plasmid sequences integrated into the host genome sequences (Fig. 4A). Within this set, it was observed that 50% of the total antibiotic resistance genes (ARGs) validated in our database were present (Fig. 5). All the identified resistance genes had an identity of more than 90% with the reference sequence (Supplementary Table 5). The *tet32* gene sequence was associated with bin_14, which belongs to the *Lachnospiraceae* family.

Virus found in the rumen

A total of 11 high-quality metagenome-assembled viral genomes (vMAGs) were identified and characterized as viral sequences (as shown in Supplementary Table 6). They had completeness rates of 92–100% and a length of 35 to 126 kbp. Another 22 MAGs were obtained with medium quality, of which 13 had completeness rates of 71–89% and a length of 25–90 kbp, and nine had completeness rates of 68–50% and lengths of 20 and 130 kbp. Any genomes with less than 50% completeness were considered low quality. All of the assembled genomes were determined to be free of contamination (Supplementary Table 6).

The majority of vMAGs had their taxonomy associated with genomes that have not yet been assigned to the ICTV phage database²⁹. There were vMAGs associated with the same viral cluster (VC); however, these do not yet have taxonomic assignment in the database: vMAG_4 and vMAG_35 were associated with the viral cluster VC_146, as well as vMAG_5 and vMAGs_18, with VC_147 and vMAG_25, and vMAG_56, with VC_200 (Supplementary Table 7).

Among the viruses that could be assigned to known genomes, vMAG_1 and vMAG_48 were assigned to the *Myoviridae* family VC_155, vMAG_40, vMAG_46 and vMAG_57 were assigned to the *Siphoviridae* family, VC_102, vMAG_7, vMAG_16, 29 and vMAG_52 were also assigned to the *Siphoviridae* family, but with genomes associated with *Streptomyces phage*, VC_87 (Fig. 6).

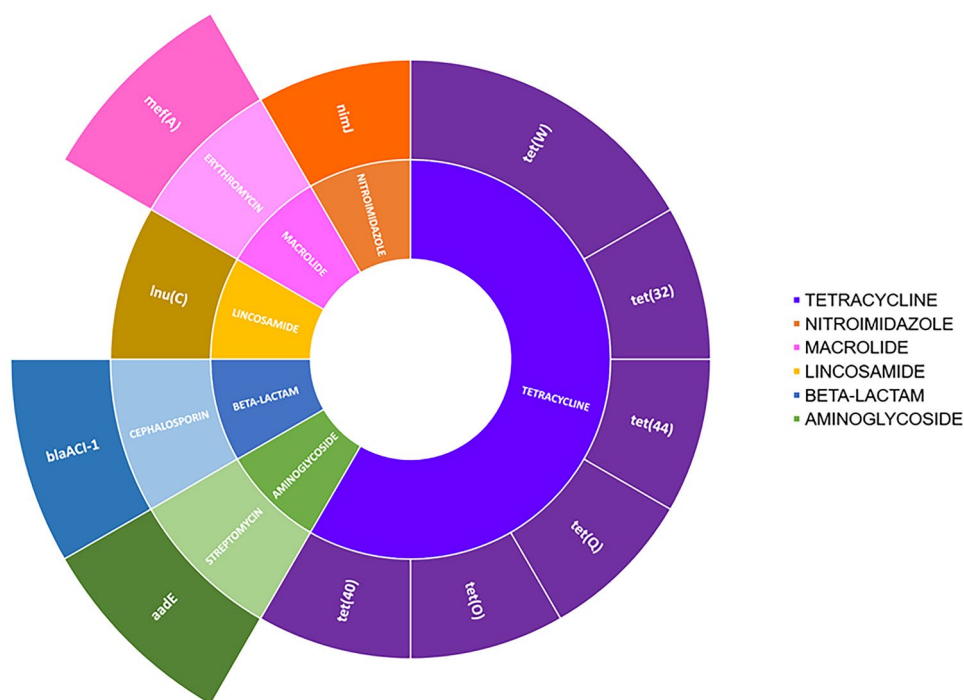


Figure 5. Antimicrobial resistance genes—ARGs. The information is presented hierarchically, starting from the center to the edges, beginning with the classes of antibiotics, followed by subclasses, and finally the resistance gene symbols.

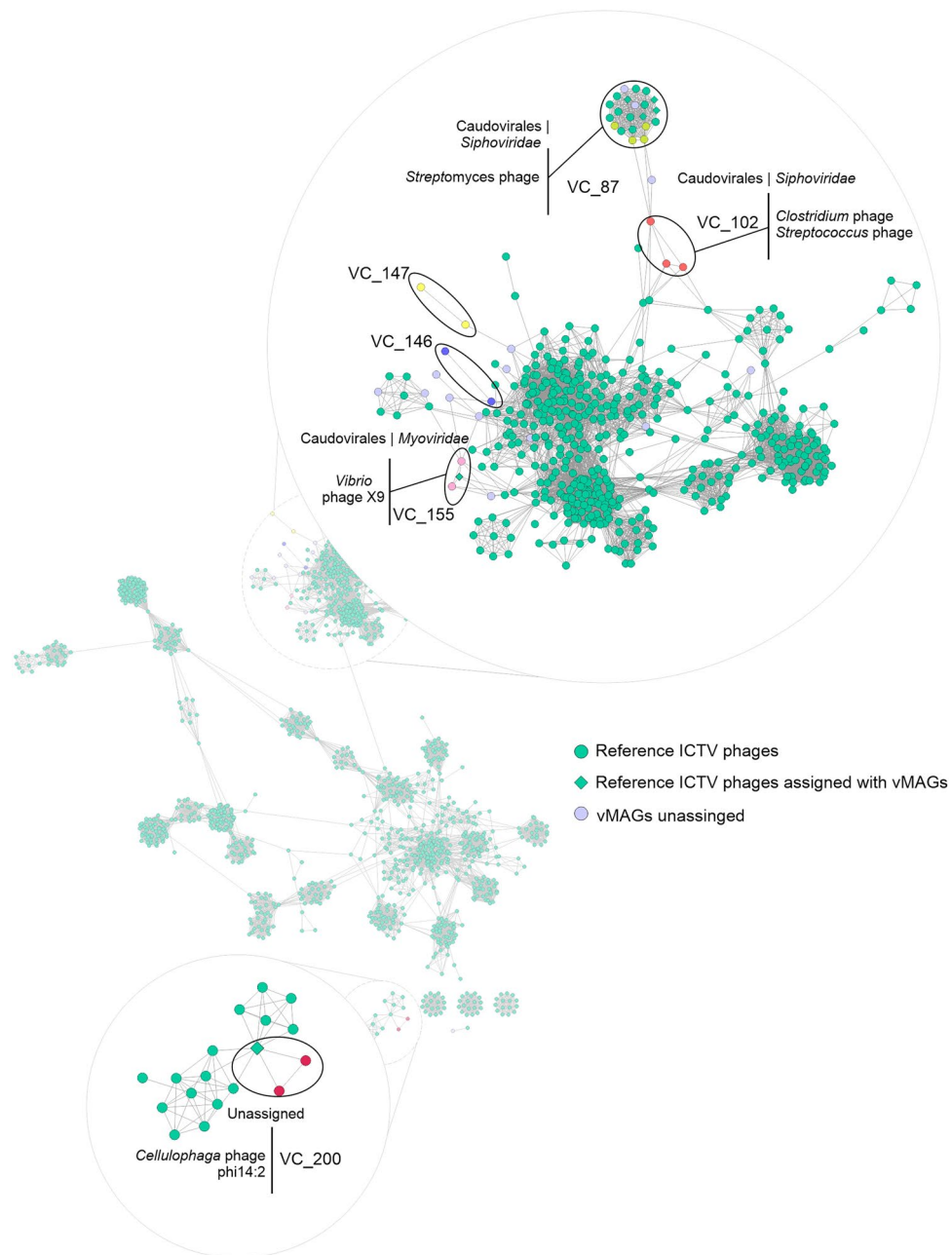


Figure 6. Taxonomic assignment of rumen viruses. Network of viral clusters assigned by vCONTACT2. Nodes with the same color have the same viral cluster, diamonds are clusters with known taxonomy and the nodes linked to them have the same viral cluster and can be assigned the same taxonomy, lilac nodes represent vMAGs that have not been assigned to clusters known.

Virus–host association

The viral clusters associated with the bacterial genomes determined by Hi-C mostly exhibited a lytic cycle. The vMAG_54, associated with the host *UBA2868 sp003535955* (*Lachnospiraceae*), presented high quality, achieved 100% completeness in its genome, and was characterized by a lytic lifestyle (Table 1).

Two co-infection events were identified, in which bacterial cells presented the assignment of more than one bacteriophage simultaneously, that is, different bacteriophages infecting the same bacterial host. Bacteriophages corresponding to vMAGs 18 and 45 infected the same bacterial cell corresponding to bin_28, and the same occurred with vMAGs 12 and 54, which infected the same bacterial cell corresponding to bin_30 (Fig. 7).

Of the 17 bins identified to be associated with viruses (Supplementary Table 12), 8 belonged to the *Lachnospiraceae* family. The bin_30 and bin_41 belong to the same genus *UBA2868*, and the other bins assigned to this family were: (bin_7) *NK4A144 sp902783395*, (bin_12) *UBA2868 sp002368575*, (bin_16) *RUG12461 sp016286115*, (bin_22) *UBA3766 sp902803015*, (bin_27) *Butyrivibrio sp017620235*, and (bin_59) *RUG191 sp002373675*. The

| Name cluster host | Host | Name cluster (phage) | Life cycle | CheckV quality | Completeness (phages) |
|-------------------|--|----------------------|------------|----------------|-----------------------|
| bin_30 | UBA2868 sp003535955 (<i>Lachno-spiraceae</i>) | vMAG_54 | Lytic | High-quality | 100.0 |
| bin_41 | g_UBA2868 (<i>Lachnospiraceae</i>) | vMAG_13 | – | Medium-quality | 87.8 |
| bin_12 | UBA2868 sp002368575 (<i>Lachno-spiraceae</i>) | vMAG_4 | Lytic | Medium-quality | 85.14 |
| bin_28 | Saccharofermentans sp003543635 | vMAG_45 | Lytic | Medium-quality | 72.67 |
| bin_22 | UBA3766 sp902803015 (<i>Lachno-spiraceae</i>) | vMAG_35 | Lytic | Medium-quality | 60.16 |
| bin_30 | UBA2868 sp003535955 (<i>Lachno-spiraceae</i>) | vMAG_12 | Lytic | Medium-quality | 54.18 |
| bin_28 | Saccharofermentans sp003543635 | vMAG_18 | Lytic | Medium-quality | 50.67 |
| bin_16 | RUG12461 sp016286115 (<i>Lachno-spiraceae</i>) | vMAG_22 | Lytic | Low-quality | 46.05 |
| bin_102 | Non-classified | vMAG_43 | Lytic | Low-quality | 39.81 |
| bin_27 | Butyrivibrio sp017620235 (<i>Lachno-spiraceae</i>) | vMAG_17 | Lytic | Low-quality | 39.28 |
| bin_58 | g_F23-D06 | vMAG_51 | – | Low-quality | 37.32 |
| bin_74 | g_UBA1711 (P3) | vMAG_42 | Lytic | Low-quality | 34.54 |
| bin_73 | Non-classified | vMAG_27 | – | Low-quality | 25.64 |
| bin_49 | Non-classified | vMAG_6 | – | Low-quality | 24.02 |
| bin_59 | RUG191 sp002373675 (<i>Lachno-spiraceae</i>) | vMAG_44 | Lytic | Low-quality | 23.57 |
| bin_7 | NK4A144 sp902783395 (<i>Lachno-spiraceae</i>) | vMAG_24 | Lytic | Low-quality | 8.71 |
| bin_26 | Saccharofermentans sp902782235 | vMAG_49 | Lytic | Not-determined | NA |

Table 1. Viruses associated with bacterial hosts by Hi-C. g_(genus).

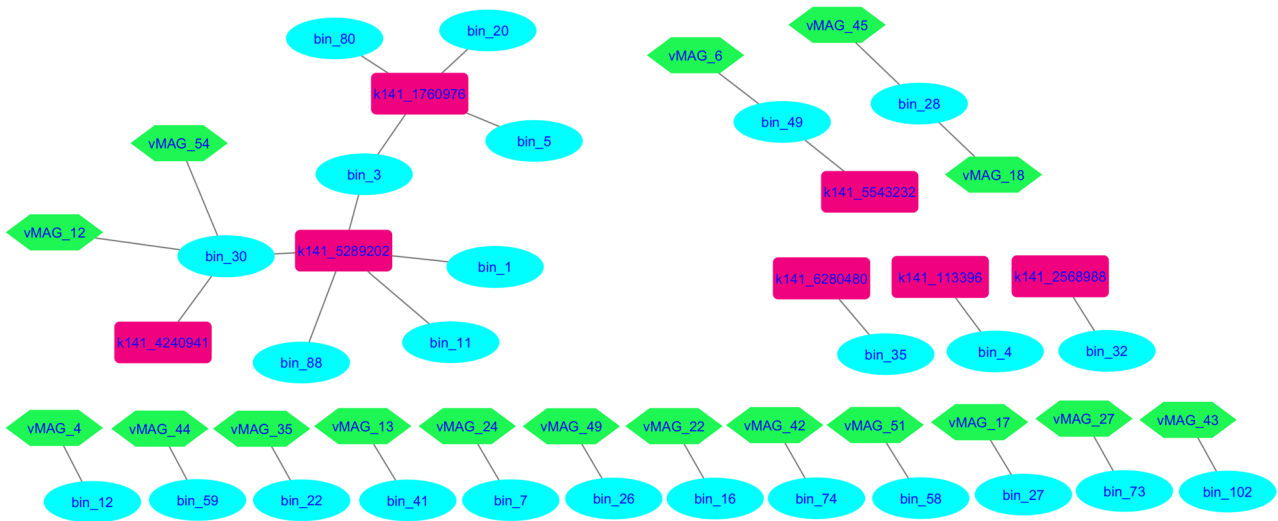


Figure 7. Association between bacterial hosts and their Hi-C-linked mobile genetic elements. Ellipses represent bacterial hosts, hexagons represent bacteriophages, and rectangles represent plasmids.

bin_26 and bin_28 were assigned to *Saccharofermentans* sp902782235 and sp003543635, respectively, and bin_58 was assigned to genus *F23-D06* and bin_74 to genus *UBA1711*, both belonging to the order *Bacteroidales*.

Discussion

Metagenomics has enabled the efficient identification of microorganisms, enzymes, and metabolic pathways that play a role in plant breakdown in the bovine rumen³⁰. This tool has also proven to be crucial for identifying the microbiota and characterizing important traits such as methane production and feed efficiency^{3,8,31}. This study is the first to explore rumen mobile genetic elements of Nellore cattle fed a grass diet by associating viruses and plasmids with their hosts through physical links between DNA molecules from the same genome using the Hi-C method²². We identified 31 links between bacteria and mobile genetic elements, co-infection events and

genes conferring antibiotic resistance. Additionally, we assembled a collection of 34 bacterial genomes with completeness > 50%, which mostly encoded pathways for central carbon and other carbohydrate metabolisms.

In this study, we identified 52 plasmid sequences, 14 of which were linked to their hosts using Hi-C. Similarly, in their evaluation of the canine fecal metagenome, Cuscó et al.²⁴ found that six Hi-C plasmids were ligated into their bacterial hosts, including five circular plasmids and a plasmid carrying the *linA* resistance gene. In our investigation, we uncovered two plasmids, k141_1760976 and k141_5289202, which are associated with multiple host clusters. As previously reported by Mo et al.³² and Stewart et al.²⁸, sequences may be shared among various bacterial species found in the rumen. Through the Hi-C approach, Stalder et al.³³ found that plasmids are efficient vectors for horizontal gene transfer and that this method is useful for tracing microorganisms that harbor antibiotic-resistant genes. We identified 12 genes conferring antibiotic resistance, and among these, six genes were found to confer resistance to tetracycline: *tet32*, *tet40*, *tet44*, *tetO*, *tetQ*, and *tetW*. These findings are consistent with a previous study by Jing and Yan³⁴, in which tetracycline resistance genes (*tet44*, *tetQ*, and *tetW*) were the most prevalent in rumen content samples and were the most abundant in the genome of *Prevotellaceae*, but *Lachnospiraceae* were also present. Furthermore, *tetQ*, *tetW*, and *tet40* were also detected in most samples derived from the rumen of 48 beef cattle from three different taurine breeds, the expression of these genes was associated with the stability of the microbiota, which may give it greater resistance to external perturbations such as diet, for example³⁵.

In addition to the genes that confer resistance to tetracycline, other classes of antibiotics, such as nitroimidazole *nimJ*, macrolide *mefA*, lincosamide *lnuC*, beta-lactam *blaACI-1*, and aminoglycoside *aadE*, were also identified. Genes that confer resistance to macrolides, lincosamide, and aminoglycosides were the most prevalent among ARGs in the study conducted by Ma et al.³⁵, with the *mefA* gene being one of the most common. Additionally, ARGs associated with plasmids were identified, indicating that plasmids should be the focus of future investigations on antimicrobial resistance in livestock. Auffret et al.³⁶ also found an increase in ARGs in rumen content samples from cattle fed diets with higher concentrate content, with an abundance of genes related to resistance to macrolides and beta-lactams, suggesting that this increase is associated with dysbiosis caused by more concentrated diets. The findings of this study are in agreement with those of an earlier in silico study conducted by Sabino et al.³⁷, who analyzed the rumen resistome in 435 microbial genomes and found a high prevalence of antibiotic resistance genes (ARGs) that provide resistance to β -lactams, glycopeptides, tetracyclines, and aminoglycosides. Recent metagenomic studies have indicated that bovine rumen is a significant contributor to antibiotic resistance³⁴, it is not necessarily related to the administration of antibiotics, considering that the rumen contents of young animals that did not receive antibiotic treatment demonstrated the presence of ARGs^{38,39}.

Advancements in bioinformatics have contributed to a deeper understanding and better characterization of the rumen virome. Yan et al.¹¹ identified 397,180 viral operational taxonomic units (vOTUs) from 975 metagenomes of 13 ruminant species, with the majority of metagenomes derived from taurine animals. Only 23 metagenomes were from *Bos indicus* raised in Kenya, and none were from Brazil, supporting the significance of this research. The viral metagenomic assemblages discovered in this study were mostly assigned to unclassified genomes, which is probably because of the limited rumen virome of *Bos indicus* data in the ICTV phage database. This finding is consistent with those of Sato et al.⁴⁰, who observed low numbers of viral operational taxonomic units shared with the RefSeq database for samples from the rumen of Japanese cattle. The assigned vMAGs belong to the families *Myoviridae* and *Siphoviridae*, which are commonly found in the rumen environment and constitute the most abundant group, along with *Mimiviridae* and *Podoviridae*^{13,16,40,41}.

A group of bacteria classified as UBA2868 sp003535955 (unclassified *Lachnospiraceae* bacterium, NCBI) was found to have the highest percentage of complete vMAGs (100%) and was characterized as high and medium quality. This genus was previously identified in a community of uncultured microorganisms from the intestines of pigs through sequencing of fecal samples^{42,43}. Bacteria from the unclassified family *Lachnospiraceae* are prevalent in the ruminal environment and serve as the central microbiome in cattle⁹. This allowed us to infer that population-associated viruses play a crucial role in the bovine rumen, rather than populations present in low abundance, which remains a challenge²¹.

Friedersdorff et al.⁴⁴ successfully isolated and sequenced active lytic phages belonging to the *Siphoviridae* family that infect *Butyrivibrio fibrisolvens*, which was found in both the rumen and feces of cattle and sheep. We discovered a lytic phage in *Butyrivibrio* sp017620235, which is commonly found in the rumen and promotes the degradation of lignocellulose and fermented carbohydrates into butyrate, formate, lactate, and acetate^{45–47}. Two bacterial species classified as Rumen Uncultured Genomes (RUGs), (bin_16) RUG12461 sp016286115 and (bin_59) RUG191 sp002373675, were also associated with viruses having a lytic life cycle.

The study's assembly bacterial genomes achieved a high completeness (> 95%) for members of the *Treponema* genus, which are known as spirochetes and have been positively correlated with feed conversion in cattle³. These species are considered to be one of the main fiber degraders in the rumen of Gir cattle, along with other bacteria, such as *Clostridium*, *Ruminococcus*, *Eubacterium*, *Butyrivibrio*, *Roseburia*, *Caldicellulosiruptor*, and *Rhodospirillum*⁴⁸. The bins identified were primarily assigned to families, such as *Lachnospiraceae*, *Bacteroidaceae*, *P3*, *Ruminococcaceae*, *Saccharofermentanaceae*, and *Treponemataceae*. These bins mainly encoded pathways related to central carbon metabolism and metabolism of various carbohydrates. One notable pathway found between these bins is the phosphate acetyltransferase-acetate kinase pathway, which is activated at high acetate concentrations, requires less ATP, and plays a role in energy metabolism⁴⁹. Additionally, the aromatic amino acid metabolism pathway with tryptophan biosynthesis was abundant among the bins in the current study. This pathway occurs only in microorganisms and plants, and tryptophan is obtained by animals via symbiosis for protein synthesis⁵⁰.

The identification of CAZymes in the bacterial genomes revealed the presence of genes encoding glycoside hydrolases and polysaccharide lyase groups, with GH2, GH3, GH5, GH13, and GH43 being the most prevalent. These groups degrade xylan polysaccharides, which are heterogeneous and require different catalytic enzymes¹⁰. In a study by Wang et al.⁵¹, the same GH families, GH2, GH3, GH13, and GH43, were found to comprise the

majority of CAZymes present in the rumen of dairy cows fed different proportions of roughage and concentrate. The GH3 group was particularly more abundant in the high-roughage diet, and these observations are consistent with the current study's findings for pasture-fed Nellore cows, where glycoside hydrolase groups were the most prevalent.

In summary, this study provides a thorough examination of the MGEs in the rumen of Nellore cattle, shedding new light on the connection between these elements and their microbial inhabitants. The impact and function of these MGEs in beef cattle are currently being investigated by our research group, we recognize the need to validate the results of the present study in a significant number of animals to better understand the ruminal dynamics of Nellore cattle kept at pasture.

Material and methods
Animal characterization and sample collection

Samples of ruminal content were collected from four Nellore (*Bos taurus indicus*) nulliparous cows cannulated in the rumen, with an average weight of (600 ± 50 kg) and age between 6 and 9 years, from the Ruminant Nutrition Laboratory, Faculty of Veterinary Medicine and Animal Science, FMVZ, located on Pirassununga (SP), in Brazil. The experimental protocols were approved by Ethics Committee on the Use of Animals of Faculty of Veterinary Medicine and Animal Science (CEUA/FMVZ) under number 7333211118. The experimental protocols were conducted in accordance with the relevant guidelines and regulations. The study design and analysis conform to the ARRIVE recommendations for animal research. The animals received a diet exclusively based on *Brachiaria brizantha* pasture and mineral salt ad libitum, which was used only for animal maintenance (Table 2) for approximately four months before the ruminal content collection. Samples were kept on ice during collection until processing and were subsequently stored at – 80 °C.

DNA extraction and sample preparation for Hi-C

DNA extraction was performed using a QIAmp Fast DNA Stool Mini Kit (Qiagen, USA). The crosslinking procedure was performed, in which the samples were processed according to the protocol recommended by the company Phase Genomics (Seattle, WA, USA) adapted by Burton et al.⁵². Briefly, the sample containing liquid and fiber (5 ml) were suspended in 1% formaldehyde (45 ml) and incubated (20 min) at room temperature with periodic shaking. Subsequently, glycine (1 g/100 ml) was added, which was then incubated (15 min) at room temperature with periodic shaking. The sample was centrifuged (1000×g for 1 min) and washed with PBS. It was then centrifuged again and the supernatant was removed. The pellets obtained formed a pool and were transferred to a cryotube (2 ml), stored at – 80 °C until sent along with the DNA extracted from the same sample, to be processed and sequenced by Phase Genomics.

Hi-C library preparation and short-read sequencing

To prepare Hi-C libraries, the Phase Genomics ProxiMeta Hi-C kit v4.0 was used as described by the manufacturer⁵³. After going through the crosslinking process, the pellets underwent lysis and their DNA–protein complexes were digested by the restriction enzymes (Sau3AI and MluCI)³², creating free ends on the DNA strands that had labeled nucleotides with biotin to create chimeric molecules composed of fragments from different physically close genome regions in vivo. Proximity-linked DNA molecules were drawn down using streptavidin beads and processed into an Illumina-compatible sequencing library.

Separately, DNA was extracted from an aliquot of the original sample, and a shotgun metagenomic library was prepared using ProxiMeta library preparation reagents. Sequencing was performed using Illumina NovaSeq, generating pairs of PE150 reads for the Hi-C and shotgun libraries. Hi-C and shotgun metagenomic sequencing files were uploaded to the Phase Genomics cloud-based bioinformatics portal for subsequent analyses.

Clustering, quality assessment, taxonomic and functional assignment of assembled microorganisms

Shotgun-derived reads were filtered and adjusted for quality, normalized using fastp⁵⁴, and assembled with MEGAHIT⁵⁵ using default options. Pairs of Hi-C reads were aligned to each assembly using BWA-MEM⁵⁶ with the -5SP options specified and all other default options to disable attempts to pair the reads according to the

| Diet | Content |
|-----------------------------|------------|
| <i>Brachiaria brizantha</i> | Ad libitum |
| Chemical composition (% DM) | |
| Dry matter | 34.1 |
| Ether extract | 1.9 |
| Crude protein | 6.9 |
| Neutral detergent fiber | 70.5 |
| Mineral salt ¹ | Ad libitum |

Table 2. Chemical composition of the diet. ¹Composition of mineral salt per kilogram of product: 160 g of Ca, 80 g of P, 30 g of S, 155 g of Na, 100 mg of Co, 1.250 mg of Cu, 62.5 mg of I, 1.125 mg of Mn, 25 mg of Se, 3.750 mg of Z.

normal Illumina settings. The SAMBLASTER program⁵⁷ was used to flag PCR duplicates that were subsequently excluded from the analysis. The alignments were then filtered with SAMtools⁵⁸ using the -F 2304 filter flag to remove nonprimary and secondary alignments.

Deconvolution, a method that uses the intracellular proximity signal captured by Hi-C as an indicator of the cellular origin of metagenome sequences, was performed using ProxiMeta^{28,32} to generate putative genomes and clusters of genome fragments. Prokaryotic clusters were assessed for quality using CheckM⁵⁹ and received preliminary taxonomic classifications using Mash, a generalized and automated method that identifies reference genomes compatible with the NCBI RefSeq database⁶⁰ (Supplementary Table 1). Subsequently, the bins were submitted to the Genome Taxonomy Database (GTDB)⁶¹ for taxonomic assignment (Supplementary Table 2).

The functional potential of the bins was evaluated by METABOLIC (METabolic and Biogeochemistry analyses in microCrobies)⁶², which detected the presence or absence of KEGG modules for the most abundant central pathways present in the genome of the microorganisms evaluated, as well as how they attributed the CAZymes and the main biogeochemical cycles present.

Mobile genetic elements

Viral contigs were annotated and used for lifestyle and viral protein annotation per contig using VIBRANT⁶³. The contigs were grouped into assembled genomes using ProxiPhage viral binning. Sequence integrity was validated with CheckV, viral genomes were classified into quality levels based on AAI or HMM estimation, high confidence ($\geq 90\%$ complete), medium confidence (80–90% complete) or low confidence ($< 80\%$ complete)^{23,64}. For taxonomic assignment, the protein sequences of the viral clusters (vMAGs) were predicted using Prodigal⁶⁵, and these sequences were used to build protein cluster profiles that were compared with sequences present in the ICTV phage database, generating networks of similarities between viral clusters with the using vCONTACT2⁶⁶.

To identify sequences with plasmid characteristics, a BLAST (Basic Local Alignment Search Tool) alignment was performed against the NCBI plasmid database³². Sequences annotated as plasmids were associated with the host using a host-finding algorithm²³. Microbial resistance genes, in turn, were identified with AMRFinderPlus, which allows the assessment of Antimicrobial Resistance gene content in the NCBI database⁶⁷.

Assignment of mobile genetic elements to their hosts

The step of assigning mobile genetic elements to their hosts was composed of several filtering criteria, with the first phase of filtering restricting connections with less than 2 Hi-C read links. To remove false positives, a connectivity rate (≥ 0.1) was used for the number of copies of mobile elements per cell (for repeated vMAGs the highest number of counts per cell was considered) and host intra-MAG connectivity (≥ 10 links). Finally, a receiver operating characteristic (ROC) curve was performed, which determines the ideal copy count cutoff value (≥ 0.1). This value is expected to be close to 1, as values close to or above 1 are considered reliable for attributing a real host²³. In cases where the optimal copy count threshold is excessively low (< 0.1), it is replaced with 0.1. This occurred in the present study, where the calculated value was 0.01.

Data availability

The data generated from the present study will be available in the Sequence Read Archive (SRA) database <https://dataview.ncbi.nlm.nih.gov/object/PRJNA1054691?reviewer=vd7njv2i8dqvoj4d6tmt82kr1t>.

Received: 13 December 2023; Accepted: 4 June 2024

Published online: 06 June 2024

References

- González-Recio, O. *et al.* Invited review: Novel methods and perspectives for modulating the rumen microbiome through selective breeding as a means to improve complex traits: Implications for methane emissions in cattle. *Livest. Sci.* **269**, 105171 (2023).
- Nkrumah, J. D. *et al.* Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle. *J. Anim. Sci.* **84**, 145–153 (2006).
- Auffret, M. D. *et al.* Identification of microbial genetic capacities and potential mechanisms within the rumen microbiome explaining differences in beef cattle feed efficiency. *Front. Microbiol.* **11**, 1229 (2020).
- McGovern, E. *et al.* Investigation into the effect of divergent feed efficiency phenotype on the bovine rumen microbiota across diet and breed. *Sci. Rep.* **10**, 1–11 (2020).
- Liu, Y. *et al.* Rumen microbiome and metabolome of high and low residual feed intake angus heifers. *Front. Vet. Sci.* **9**, 812861 (2022).
- Roehe, R. *et al.* Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* **12**, e1005846 (2016).
- Zhang, Q. *et al.* Bayesian modeling reveals host genetics associated with rumen microbiota jointly influence methane emission in dairy cows. *ISME J.* **14**, 2019–2033 (2020).
- Martínez-Álvarez, M. *et al.* Bovine host genome acts on rumen microbiome function linked to methane emissions. *Commun. Biol.* **5**, 1–16 (2022).
- 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).
- Gharechahi, J. *et al.* Lignocellulose degradation by rumen bacterial communities: New insights from metagenome analyses. *Environ. Res.* **229**, 115925 (2023).
- Yan, M. *et al.* Interrogating the viral dark matter of the rumen ecosystem with a global virome database. *Nat. Commun.* **14**, 1–16 (2023).
- Luo, X. Q. *et al.* Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. *Microbiome* **10**, 1–18 (2022).
- Anderson, C. L., Sullivan, M. B. & Fernando, S. C. Dietary energy drives the dynamic response of bovine rumen viral communities. *Microbiome* **5**, 1–19 (2017).
- Kav, A. B. *et al.* Insights into the bovine rumen plasmidome. *Proc. Natl. Acad. Sci. USA* **109**, 5452–5457 (2012).

15. Lobo, R. R. & Faciola, A. P. Ruminal phages—A review. *Front. Microbiol.* **12**, 763416 (2021).
16. Solden, L. M. *et al.* Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. *Nat. Microbiol.* **3**, 1274–1284 (2018).
17. Orpin, C. G. & Munn, E. A. The occurrence of bacteriophages in the rumen and their influence on rumen bacterial populations. *Experientia* **30**, 1018–1020 (1974).
18. Altermann, E., Schofield, L. R., Ronimus, R. S., Beatty, A. K. & Reilly, K. Inhibition of Rumen methanogens by a novel archaeal lytic enzyme displayed on tailored bionanoparticles. *Front. Microbiol.* **9**, 2378 (2018).
19. Park, S. Y. *et al.* Characterization of two lytic bacteriophages, infecting *Streptococcus bovis/equinus* complex (SBSEC) from Korean ruminant. *Sci. Rep.* **13**, 1–16 (2023).
20. Ross, E. M., Petrovski, S., Moate, P. J. & Hayes, B. J. Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. *BMC Microbiol.* **13**, 1–11 (2013).
21. Bickhart, D. M. *et al.* Assignment of virus and antimicrobial resistance genes to microbial hosts in a complex microbial community by combined long-read assembly and proximity ligation. *Genome Biol.* **20**, 1–18 (2019).
22. Marbouty, M., Baudry, L., Cournac, A. & Koszul, R. Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. *Sci. Adv.* **3**, e1602105 (2017).
23. Uritskiy, G. *et al.* Accurate viral genome reconstruction and host assignment with proximity-ligation sequencing. *bioRxiv* <https://doi.org/10.1101/2021.06.14.448389> (2021).
24. Cuscó, A., Pérez, D., Viñes, J., Fàbregas, N. & Francino, O. Novel canine high-quality metagenome-assembled genomes, prophages and host-associated plasmids provided by long-read metagenomics together with Hi-C proximity ligation. *Microb. Genom.* **8**, 802 (2022).
25. Braun, V. & Killmann, H. Bacterial solutions to the iron-supply problem. *Trends Biochem. Sci.* **24**, 104–109 (1999).
26. Canfield, D. E., Kristensen, E. & Thamdrup, B. The iron and manganese cycles. *Adv. Mar. Biol.* **48**, 269–312 (2005).
27. Bohra, V., Dafale, N. A. & Purohit, H. J. Understanding the alteration in rumen microbiome and CAZymes profile with diet and host through comparative metagenomic approach. *Arch. Microbiol.* **201**, 1385–1397 (2019).
28. Stewart, R. D. *et al.* Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* **9**, 870–870 (2018).
29. Krupovic, M. *et al.* Bacterial viruses subcommittee and archaeal viruses subcommittee of the ICTV: Update of taxonomy changes in 2021. *Arch. Virol.* **166**, 3239–3244 (2021).
30. Hua, D., Hendriks, W. H., Xiong, B. & Pellikaan, W. F. Starch and cellulose degradation in the rumen and applications of metagenomics on ruminal microorganisms. *Animals* **12**, 3020 (2022).
31. Wallace, R. J. *et al.* The rumen microbial metagenome associated with high methane production in cattle. *BMC Genom.* **16**, 1–14 (2015).
32. Mo, P. *et al.* Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. *bioRxiv* <https://doi.org/10.1101/198713> (2017).
33. Stalder, T., Press, M. O., Sullivan, S., Liachko, I. & Top, E. M. Linking the resistome and plasmidome to the microbiome. *ISME J.* **13**, 2437–2446 (2019).
34. Jing, R. & Yan, Y. Metagenomic analysis reveals antibiotic resistance genes in the bovine rumen. *Microb. Pathog.* **149**, 104350 (2020).
35. Ma, T. *et al.* Expressions of resistome is linked to the key functions and stability of active rumen microbiome. *Anim. Microbiome* **4**, 1–17 (2022).
36. Auffret, M. D. *et al.* The rumen microbiome as a reservoir of antimicrobial resistance and pathogenicity genes is directly affected by diet in beef cattle. *Microbiome* **5**, 1–11 (2017).
37. Sabino, Y. N. V. *et al.* Characterization of antibiotic resistance genes in the species of the rumen microbiota. *Nat. Commun.* **10**, 1–11 (2019).
38. Chai, J., Zhuang, Y., Cui, K., Bi, Y. & Zhang, N. Metagenomics reveals the temporal dynamics of the rumen resistome and microbiome in goat kids. *Microbiome* **12**, 1–18 (2024).
39. Zhuang, Y. *et al.* Altered microbiota, antimicrobial resistance genes, and functional enzyme profiles in the rumen of yak calves fed with milk replacer. *Microbiol. Spectr.* **12**, e0131423 (2024).
40. Sato, Y. *et al.* Characteristics of the rumen virome in Japanese cattle. *bioRxiv* <https://doi.org/10.1101/2023.03.20.532305> (2023).
41. 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).
42. 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).
43. Crossfield, M. *et al.* Archaeal and bacterial metagenome-assembled genome sequences derived from pig feces. *Microbiol. Resour. Announc.* **11**, e0114221 (2022).
44. Friedersdorff, J. C. A. *et al.* The isolation and genome sequencing of five novel bacteriophages from the rumen active against *Butyrivibrio fibrisolvens*. *Front. Microbiol.* **11**, 522243 (2020).
45. Kopečný, J., Zorec, M., Mrázek, J., Kobayashi, Y. & Marínšek-Logar, R. *Butyrivibrio hungatei* sp. nov. and *Pseudobutyrvibrio xylanivorans* sp. nov., butyrate-producing bacteria from the rumen. *Int. J. Syst. Evol. Microbiol.* **53**, 201–209 (2003).
46. Moon, C. D. *et al.* Reclassification of *Clostridium proteoclasticum* as *Butyrivibrio proteoclasticus* comb. nov., a butyrate-producing ruminal bacterium. *Int. J. Syst. Evol. Microbiol.* **58**, 2041–2045 (2008).
47. Palevich, N. *et al.* Complete genome sequence of the polysaccharide-degrading rumen bacterium *Pseudobutyrvibrio xylanivorans* MA3014 reveals an incomplete glycolytic pathway. *Genome Biol. Evol.* **12**, 1566–1572 (2020).
48. Pandit, R. J. *et al.* Microbiota composition, gene pool and its expression in Gir cattle (*Bos indicus*) rumen under different forage diets using metagenomic and metatranscriptomic approaches. *Syst. Appl. Microbiol.* **41**, 374–385 (2018).
49. Zhang, S. *et al.* Metabolic engineering for efficient supply of acetyl-CoA from different carbon sources in *Escherichia coli*. *Microb. Cell Fact.* **18**, 1–11 (2019).
50. Parthasarathy, A. *et al.* A Three-Ring circus: Metabolism of the three proteogenic aromatic amino acids and their role in the health of plants and animals. *Front. Mol. Biosci.* **5**, 342220 (2018).
51. Wang, L., Zhang, G., Xu, H., Xin, H. & Zhang, Y. Metagenomic analyses of microbial and carbohydrate-active enzymes in the rumen of holstein cows fed different forage-to-concentrate ratios. *Front. Microbiol.* **10**, 441658 (2019).
52. Burton, J. N., Liachko, I., Dunham, M. J. & Shendure, J. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3: Genes Genomes Genetics* **4**, 1339–1346 (2014).
53. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
54. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
55. Li, D. *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).
56. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
57. Faust, G. G. & Hall, I. M. SAMBLASTER: Fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
58. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078 (2009).

59. 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 (2015).
60. Ondov, B. D. *et al.* Mash Screen: High-throughput sequence containment estimation for genome discovery. *Genome Biol.* **20**, 1–13 (2019).
61. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
62. Zhou, Z. *et al.* METABOLIC: High-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* **10**, 1–22 (2022).
63. 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* **8**, 1–23 (2020).
64. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2020).
65. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 1–11 (2010).
66. Bin Jang, H. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
67. Feldgarden, M. *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 1–9 (2021).

Acknowledgements

The authors express their gratitude to the funding agencies that facilitated this research. The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grants 2022/05541-9, 2022/09378-5 and 17/20084-5 and the Fundação de Estudos Agrários Luiz de Queiroz (FEALQ).

Author contributions

C.A.F. collected the samples, performed the laboratory procedures until sending them to Phase Genomics for sequencing, analyzed the data obtained, compiled the results and wrote the initial manuscript. A.T.N. collaborated with sampling and writing the manuscript. M.D.P. contributed to data analysis and preparation of figures. E.C.M., O.S.G. and P.A.A. contributed with bioinformatics analyses. F.P.J. and P.H.M.R. collected the samples and provided the animal data. H.F. was the project supervisor and reviewed the initial manuscript. All authors reviewed the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-63951-7>.

Correspondence and requests for materials should be addressed to H.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024