Animal Nutriomics

cambridge.org/anr

Rumen microbiome nutriomics: Harnessing omics technologies for enhanced understanding of rumen microbiome functions and ruminant nutrition

Zhongtang Yu[1], Ming Yan[1] and Jiakun Wang[2]

[1]Department of Animal Sciences and Center of Microbiome Science, The Ohio State University, Columbus, OH, USA and [2]Institute of Dairy Sciences, College of Animal Sciences, Zhejiang University, Hangzhou, China

Review

Cite this article: Yu Z, Yan M, Wang J (2024) Rumen microbiome nutriomics: Harnessing omics technologies for enhanced understanding of rumen microbiome functions and ruminant nutrition. *Animal Nutriomics* **1**, e10, 1–11. https://doi.org/10.1017/anr.2024.10

Received: 4 March 2024
Revised: 6 May 2024
Accepted: 3 June 2024

Keywords:
meta-omics; rumen microbiome; rumen microbiome nutriomics; integrated data analyses

Corresponding author: Zhongtang Yu;
Email: yu.226@osu.edu

Abstract

The rumen microbiome has attracted tremendous interest among microbiologists and ruminant nutritionists because of its crucial role in mediating feed digestion and fermentation and supplying most of the energy, nutrients, and precursors for producing ruminant products. The application of various omics technologies, including metataxonomics, metagenomics, metatranscriptomics, metaproteomics, and metabolomics, have enabled unprecedented investigations into this ecosystem, shedding new light on its interactions with diet and animals and its relationships with key production traits. Despite the valuable insights these omics technologies provide, each has its unique utility and inherent limitations. Achieving a holistic characterization of the rumen microbiome and deciphering its causal relationship with diet and key animal production traits remain an ongoing endeavor. In this perspective review paper, we highlight the limitations of individual technologies and advocate for an integrated multi-omics approach and data analyses in studying the intricate relationships between diet, rumen microbes, and ruminant nutrition. This approach, termed "rumen microbiome nutriomics," aims to comprehensively understand the rumen microbiome in the context of diets and animal productivity. Our emphasis lies in recognizing the necessity of integrated analysis across multiple data layers, encompassing data of diet, rumen microbiome features, animal genotypes, and production traits and identifying the causal relationship among them. We also call for collaborative efforts to develop a comprehensive rumen microbiome genome database, including prokaryotes, protozoa, fungi, and viruses. Furthermore, standardization of processes and analyses is crucial to address the variability observed in the literature, facilitating comparison of results among future studies and enabling robust data reanalysis through advanced data analytics.

Introduction

Domesticated ruminants are a crucial source of high-quality proteins to meet human protein and nutrient requirements. They depend on the rumen microbiome to produce the primary source of energy, nutrients, and precursors for protein production. This unique microbiome features multi-kingdoms of remarkably diverse microbes, including bacteria and archaea as prokaryotes, protozoa and fungi as eukaryotes, and viruses. Bacteria are the most abundant and diverse, encompassing thousands of species (Creevey et al. 2014; Kim et al. 2011). Rumen archaea are primarily methanogens, while protozoa are nearly exclusively ciliates. Despite being less diverse and abundant, only 10^4 – 10^5 individuals per ml of rumen fluid, protozoa can match bacteria in terms of biomass (Andersen et al. 2023). Rumen fungi represent only 10 – 16% of total rRNA transcript abundance (Elekwachi et al. 2017) and less than 20% of the rumen microbial biomass (Rezaeian et al. 2004). Despite being the smallest, rumen viruses are diverse and abundant (Gilbert et al. 2020; Yan et al. 2023). These microbes form a dynamic and finely-tuned ecosystem. Their populations and metabolism can shift in response to changes in diet, allowing ruminants to adapt to different nutritional regimes.

Previous studies have provided fundamental information on the capability of rumen microbes, primarily bacteria. As the most abundant microbes, bacteria play the most crucial role in the rumen functions, such as feed digestion, fermentation, and microbial protein synthesis. Archaea produce enteric CH_4, a potent greenhouse gas that raises significant environmental concerns associated with ruminant production. Protozoa participate in feed digestion and fermentation, but as predators, they engulf microbial cells and degrade microbial protein, significantly contributing to the intraruminal recycling of microbial protein, a process primarily responsible for the lower nitrogen utilization efficiency in ruminants than in nonruminants. Fungi are not abundant but possess a unique ability and high activity to digest feed fiber (Bhagat et al. 2023). Rumen viruses do not directly digest or ferment feed. Still, by

ysing their hosts or providing auxiliary metabolic genes and other genes, they can profoundly impact the functions and metabolic activities of various rumen microbes, including those that form the core rumen microbiome, in both top-down and bottom-up manners (Yan et al. 2023). The rumen microbes constitute an intricate ecosystem by interacting with each other, the diet, and the hosts, and this ecosystem is responsible for converting feed into energy, nutrients, and precursors that ruminants can utilize. Therefore, the rumen microbiome can profoundly affect feed efficiency, animal health, productivity, quality of products (meat, milk, and wool), and the environmental footprint of the ruminant industry. Understanding the diversity, composition, and functions of the rumen microbiome and its interaction with diet and host has been a long-term pursuit of research over the past century.

Despite considerable progress in understanding the rumen microbiome, knowledge gaps remain regarding the interactions of most rumen microbes with diet and host and their contributions to animal nutrition and productivity. First, the role and significance of specific microbial species in enhancing nutrient utilization and reducing CH_4 emissions are not yet fully understood. Second, the crosstalk between different microbial taxa and their dynamic interactions with the diet and the host requires further investigation. Third, the heritability (h^2) of rumen microbes, which reflects the influence of host genotypes on shaping the rumen microbiome and its functions (Martinez Boggio et al. 2022), needs to be explored more. Fourth, the microbiability (m^2) of key animal production traits, calculated as the proportion of variance in a specific production trait explained by the rumen microbiome (Difford et al. 2018), has only started to be assessed. Fifth, comprehensive investigations into the resilience of the rumen microbiome to environmental stressors, such as heat stress, are necessary to develop sustainable livestock management practices. Finally, the exploration of the rumen virome is in its initial stages, and its influence on the populations of rumen microbes or the overall rumen functions remains to be determined

Understanding the complex relationship between diet, rumen microbiome, host, and specific production traits presents some challenges. The rumen microbiota (or microbial) composition can considerably vary even among cohorts of the same breed fed the same diet, making it challenging to attribute different production traits to differences in the rumen microbiome. Experimental design and analyses, including sequence data processing and bioinformatic analyses, lack standardization, which makes it difficult to compare results across different studies. Additionally, the microbiome data generated from metataxonomics and metagenomics are sparse, high-dimensional, zero-inflated, and compositional, necessitating complicated statistical analyses. Furthermore, correlations can be observed among diets, the rumen microbiome, rumen functions, and production traits. However, the complex and dynamic nature of this diverse microbiome poses challenges in establishing unequivocal causal relationships. Additionally, most studies identify rumen microbes only at the genus level, but species, even strains, can vary significantly in their metabolism, activity, and contributions to overall rumen functions. Furthermore, from an analytical perspective, the data layers generated by individual meta-omics technologies can exhibit interactions with animal production traits. Therefore, the integration of various omics technologies and data analyses is crucial for a comprehensive understanding of the rumen microbiome and its relationship with diet and animal nutrition. This integrated multi-omics approach, combined with integrated analysis of the multiple layers of data, is referred to as "rumen microbiome nutriomics."

Omics technologies for rumen microbiome nutriomics

Since the early to mid-2000s, omics have become the primary technologies in microbiome research, including metataxonomics, metagenomics, metatranscriptomics, metaproteomics, and metabolomics coupled with bioinformatics. These meta-omics technologies have enabled comprehensive investigations of the rumen microbiome, leading to an unrepresented understanding and appreciation of its vast diversity, composition, functional capacity, and association with diets and key animal production traits, such as feed efficiency and methane emission. However, each of these omics technologies has its inherent limitations.

Metataxonomics

The 16S rRNA gene is among the few phylogenetic markers analyzed through high-throughput sequencing in early studies profiling microbiomes, including the rumen microbiome. Metataxonomics involves PCR amplification, high-throughput sequencing of phylogenetic markers, and bioinformatic analysis to taxonomically identify the microbes within microbiomes (Denman et al. 2018). It is the first omics technology used in comprehensively profiling the rumen microbiome, greatly contributing to our understanding of its extensive diversity. Although it can help taxonomically identify most cellular rumen microbes, it has several limitations (Denman et al. 2018). First, the preparation of amplicon sequencing libraries involved PCR, but PCR introduces biases (Silverman et al. 2021) stemming from the choice of phylogenetic regions targeted and the primers used (Laursen et al. 2017; Tremblay et al. 2015; Yu and Morrison 2004). These biases can compromise differential abundance analysis (DAA), especially for the minor taxa, posing challenges in comparing DAA results among studies that use different marker regions and primers. Second, while metataxonomics can cost-effectively detect and identify most cellular microbes, the short amplicon sequences lack the necessary taxonomic resolution to support species-level classification (Johnson et al. 2019). This limitation is particularly profound in the analysis of rumen ciliates due to the highly conserved nature of their 18S rRNA gene (Somasundaram and Yu 2024). Third, although comparing the marker sequences to databases with specific tools like PICRUSt2 (Douglas et al. 2020) and CowPI (Wilkinson et al. 2018) can help predict the functional capability of the rumen microbiome, it does not provide direct evidence of its functional capacities. Additionally, the lack of species-level identification and the detection of numerous unclassified microbes constrain the depth of functional insights. Finally, metataxonomics cannot detect viruses or phages because they do not have conserved phylogenetic markers. Nevertheless, metataxonomics is still valuable in rumen microbiome studies. Sequencing alternative markers, such as the internal transcribed spacers and 23S or 28S rRNA genes, can help enhance the taxonomic resolution, particularly by sequencing the entire length of these markers. The full lengths of all the commonly used phylogenetic markers can be sequenced using synthetic long-read sequencing technologies, such as LoopSeq (Callahan et al. 2021), or long-read sequencing technologies, such as MinION (https://nanoporetech.com/) and Sequell II (https://www.pacb.com/technology/hifi-sequencing/sequel-system/), enhancing the accuracy and resolution of taxonomic assignments (Abellan-Schneyder et al. 2021). Furthermore, as demonstrated by Greengene2 (McDonald et al. 2023), amalgamating databases of phylogenetic markers and genomes can improve the utility of metataxonomics in analyzing the rumen microbiome.

Metagenomics

In brief, metagenomics encompasses shotgun sequencing and a series of bioinformatic analyses of DNA directly extractive from microbiome samples. This omics technology is commonly used to unveil the taxonomic diversity and functional capacities of microbiomes, and it has proven to be one of the most powerful omics technologies in rumen microbiome research (e.g., [Amin et al. 2022]). Through taxonomic assignments of metagenomic sequences, contigs, or metagenome-assembled genomes (MAGs), metagenomics can potentially identify all microbes, including viruses, providing insights into the overall diversity, composition, and structure. For sequence-based taxonomic assignments, several bioinformatics programs are available, such as MetaPhlAn2 (Truong et al. 2015), Kraken2 (Lu and Salzberg 2020), mOTUs2 (Milanese et al. 2019), and Kaiju (Menzel et al. 2016). Contig-based taxonomic assignment enhances classification accuracy, and several bioinformatics tools are available for this purpose, including DIAMOND (Buchfink et al. 2015), CAT (von Meijenfeldt et al. 2019), and MetaBinG2 (Mirdita et al. 2021). MAG-based taxonomic assignment further enhances taxonomic classification. Species-level taxonomic assignment can be achieved with GTDB-Tk v2 (Chaumeil et al. 2022) and its genome database and taxonomy (Parks et al. 2018). As sequencing costs decrease, metagenomic sequencing depth increases, increasing the number of high-quality MAGs (>90% complete with <5% contamination) and thus filling some of the gaps in genome databases. Improvements in reference genome databases will facilitate profile microdiversity and population dynamics at species, even strain levels. Notably, strain-level profiling of metagenomes has been demonstrated using inStrain (Olm et al. 2021) in a recent study on the interactions between rumen microbiome and virome (Yan and Yu 2024). Therefore, genome-centric and genome-resolved metagenomics will further enhance taxonomic profiling of the rumen microbiome, particularly at the species and strain levels.

Metagenomics can uncover the functional potential of the entire rumen microbiome, along with discovering novel genes, enzymes, and pathways. Indeed, early metagenomic studies revealed an incredible repertoire of various genes, shining new light on the functional diversity and potential of the rumen microbiome (Brulc et al. 2011). However, metagenomics cannot distinguish genes from dead versus viable microbes. Additionally, the "bag-of-genes" generated through gene-centric metagenomics provides scant insight into genomic architecture. This approach also has limited capacity to unveil new microbial species or reconstruct the metabolic networks (MN) of individual microbes (Frioux et al. 2020). Genome-centric and genome-resolved metagenomics can address some of the limitations by constructing MAGs and also provide opportunities to estimate the growth rates of individual prokaryotes represented by MAGs (Joseph et al. 2022; Korem et al. 2015), illuminating the population dynamics of individual microbes within rumen microbiomes (Zhang et al. 2022). Nevertheless, genome-centric and genome-resolved metagenomics face several challenges. First, metagenomic sequences are often short (<300 bp), making it challenging and computing-demanding to assemble MAGs, particularly for rumen microbes at low abundance, including ciliates and fungi, which also have large complex genomes. Metagenomic sequences from multiple samples of the same individual or treatment can be co-assembled and binned to help recover MAGs of low abundance species, but this approach leads to poor results when the samples have high intraspecies diversity, and it is computationally consuming (Delgado and Andersson 2022). Second, genome reconstruction also has biases (Nelson et al. 2020), leading to over- or under-representation of specific microbial taxa, affecting the accuracy of metagenomic analysis. Third, assigning functions to some genes in metagenomic datasets can be challenging due to gaps in reference genome databases and many unknown or hypothetical genes. Indeed, about one-third of the protein-coding genes from bacterial genomes could not be functionally annotated (Bileschi et al. 2022). Deep learning models emerge as an effective tool to enhance functional annotation (Bileschi et al. 2022).

Ongoing research efforts are focusing on addressing the above challenges. Integrating short- and long-read sequencing technologies can improve sequence assembly, increasing high-quality MAGs. Developing and refining bioinformatics tools can enhance the quality of MAGs and streamline the metagenomics process. For example, using machine learning, CheckM2 improves MAG quality assessment (Chklovski et al. 2022). Further, expanding and refining reference databases can improve the accuracy of taxonomic classification and functional annotation (Stewart et al. 2019; Xie et al. 2021; Yan and Yu 2024). Integration of metagenomics, particularly genome-centric and genome-resolved metagenomics, with other omics technologies, such as metatranscriptomics and metaproteomics, along with the continued refinement of bioinformatics tools, can provide more comprehensive insights into the rumen microbiome and its complex interactions with diets, animals, and production traits. It should also be noted that genome-centric and genome-resolved metagenomic studies have predominantly focused on rumen bacteria and archaea, thereby neglecting rumen protozoa, fungi, and viruses. Leveraging on the recent bioinformatics tools specifically developed for viral sequence analyses, such as VirSorter2 (Guo et al. 2021), VIBRANT (Kieft et al. 2020), and CheckV (Nayfach et al. 2021), several recent studies have successfully revealed that the rumen virome is highly diverse and can infect a wide range of rumen microbes, including the core rumen microbiome (Yan et al. 2023), responds to diets (Anderson et al. 2017), and associates with microbial diversification, community dynamic, and specific production traits (Yan and Yu 2024). New bioinformatics tools capable of discerning eukaryotic signals amid metagenomic sequences, coupled with newly sequenced genomes of rumen protozoa and fungi, will significantly enhance the analysis of these rumen eukaryotic microbes within rumen metagenomic datasets.

Metatranscriptomics

Through sequencing and bioinformatic analysis of RNA, metatranscriptomics reveals actively expressed genes, collectively referred to as the transcriptome. rRNA is commonly removed before conducting RNA-Seq to enhance sequencing efficiency and allow for more precise sequencing of mRNA alongside non-coding RNA and small RNA. Hence, metatranscriptomics illuminates the ongoing metabolic and other biological processes within microbiomes. This omics technology has yielded valuable insights into how the rumen microbiome responds to dietary alterations or interfaces with specific rumen functionalities and production traits at the transcriptional level. Previous metatranscriptomic investigations have focused on genes exhibiting differential expressions between diets of feed additives (e.g., (Jize et al. 2022; Pitta et al. 2022)), animal productivities (e.g., (Park et al. 2022; Xue et al. 2022)), or breeds (e.g., (Li et al. 2019; Zhang et al. 2020)). Linking the expressed genes to the specific host microbes can be challenging with such

a ene-centric metatranscriptomic approach. Furthermore, rumen metatranscriptomes have been analyzed for transcripts of prokaryotes. The eukaryotic transcripts and the genomes of RNA viruses should also be analyzed in future studies.

Genome-centric metatranscriptomics focuses on the analysis of transcriptional activity within microbiomes, potentially at the level of individual genomes. This approach employs RNA-Seq and compares transcript sequences to individual genomes or MAGs. Hence, it enables researchers to (i) associate transcripts with the expressing genomes or MAGs and (ii) reconstruct genome-scale MN or models for individual microbes. Such information facilitates a more precise evaluation of the contributions of those microbes to the critical metabolic processes, such as feed digestion and fermentation, protein synthesis, and CH_4 emissions. Additionally, differential gene expression (DGE) and pathway enrichment analyses are crucial in revealing how microbial activities respond to variations in diets and interface with rumen functions and animal productivity. Furthermore, genome-centric metatranscriptomics facilitates identifying the rumen fungi that produce microRNA-like RNAs (no evidence is available indicating that bacteria, archaea, and protozoa produce microRNAs) and all cellular rumen microbes produce small RNAs. While genome-centric metatranscriptomics can potentially provide dynamic insights into rumen functions at the genome level and dynamics in the rumen microbiome, it faces several challenges with low-abundance transcripts. Gaps in reference genome databases and the presence of unknown or hypothetical genes further hinder the identification of some expressed genes (Shakya et al. 2019). Furthermore, genome-centric metatranscriptomics can be biased toward cultured microbes with well-annotated genomes. As sequencing costs decrease and reference genome databases expand, metagenome-centric metatranscriptomics is poised to surpass gene-centric metatranscriptomics.

Metaproteomics

Metaproteomics, the study of all the proteins expressed in a microbiome, the metaproteome, offers a snapshot of the expressed proteins therein. Unlike metagenomics or metatranscriptomics, it provides a "snapshot" of actively working proteins, revealing the actual metabolic landscape at the sampling time. Pathway enrichment analysis can help identify the pathways corresponding to the identified proteins, providing dynamic insights into the activities of a microbiome. This extends beyond the capabilities of metagenomics or metatranscriptomics, furnishing a more direct perspective on the actual functional processes and their connection to animal productivity (Andersen et al. 2021). Studies have used metaproteomics in investigating the metabolic influence of rumen protozoa within the rumen microbiome (Andersen et al. 2023) and its responses to dietary interventions (Trautmann et al. 2023) and heat stress (Li et al. 2021). Metaproteomics can also help identify biomarkers associated with specific microbial functions, microbiome dysbiosis, rumen functions, or production traits. However, metaproteomics can face several challenges, as demonstrated in other microbiomes, including the complexity and diversity of the rumen microbiome, limitations in detecting low-abundance proteins, and issues with identical peptides from homologous proteins (Heyer et al. 2017; Lohmann et al. 2020; Miura and Okuda 2023). Finally, the lack of complete genomes and protein databases for many rumen microbes, particularly rumen fungi and protozoa, hinders precise annotation and taxonomic assignment, leaving some identified proteins with unknown origins. These challenges are further exacerbated by the presence of dietary proteins in the rumen. To fully harness the potential of genomic-centric metaproteomics for studying the rumen microbiome, comprehensive reference genome databases specific to this microbiome are essential. The Rumen Microbial Global Network or a similar international network can facilitate collaborative efforts to compile existing and future genomics data including MAGs. These databases, designed to minimize gaps in the representation of key rumen microbes, will enable genome-centric metaproteomics. Such an approach promises unprecedented insights into the roles of key rumen microbes and their impacts on various rumen functions and production traits. Ultimately, this information will empower efforts to optimize the rumen microbiome for improved animal health and productivity.

Metabolomics

Metabolomics leverages proton nuclear magnetic resonance (NMR) spectroscopy and gas or liquid chromatography coupled with mass spectrometry (GC-MS or LC-MS) or tandem MS (GC-MS/MS or LC-MS/MS) to separate and identify individual metabolites. Targeted metabolomics analyzes a predefined set of related metabolites, whereas untargeted metabolomics involves global metabolic profiling. Metabolomics enables the elucidation of the complex metabolic profiles within the rumen microbiome, offering valuable insights into its functional activities. Univariate and multivariate statistical analyses can help identify specific metabolites that differ between animals or treatments. Moreover, pathway enrichment analysis can help identify the metabolic pathways that are influenced or differentially expressed. Metabolomics has been used in examining how the rumen metabolic profiles respond to dietary shifts (Ali et al. 2023; Ren et al. 2023), dietary supplements (de Poppi et al. 2021; Li et al. 2022a), stresses (Feng et al. 2022; Li et al. 2023), and health status (Eom et al. 2021; Mu et al. 2022). Furthermore, metabolomics aids in the identification of rumen metabolites or pathway enrichment indicative of divergent rumen functions or production traits, including residual body weight gain (Idowu et al. 2023), RFI (Liu et al. 2022b), and efficiency of milk production in dairy cows (Xue et al. 2022).

Rumen metabolomics also faces several challenges. First, the rumen microbiome produces a myriad of metabolites at various concentrations, but only a relatively small number of them can be detected or identified. Second, identifying and annotating rumen microbiome metabolites face challenges because many metabolites lack known reference standards, leading to uncertainties in result interpretation. Third, the accurate assessment of the metabolic response of the rumen metabolome necessitates the quantification of metabolites, but the complex matrices of rumen samples may compromise the reliability of quantification. Fourth, several metabolomic databases like BMDB (www.bovinedb.ca), MetaboBank (https://metabo.ca), and MetaboLights (http://www.ebi.ac.uk/metabolights/), as well as pathway databases like KEGG (https://www.genome.jp/kegg/) and MetaCyc (https://metacyc.org/) can be used to map metabolites to their corresponding metabolic pathways. However, gaps in these databases constrain the reliable identification of metabolites and linking metabolites or metabolic pathways to the producers. The development of metabolomic databases specific to the rumen ecosystem and advancements in bioinformatics tools for metabolite annotation and pathway analysis will contribute to a more accurate and meaningful interpretation of rumen microbiome metabolomic data. Furthermore, integration with other omics technologies, such as genomics, genome-centric metagenomics, metatranscriptomics, and metaproteomics, is essential to further enhance the capability of metabolomic analysis of the rumen microbiome.

As demonstrated in a recent study (Idowu et al. 2023), integrating currently used LC-MS with other techniques, such as isotope labeling, can increase the sensitivity of metabolite detection.

Bioinformatics and databases

Bioinformatics is essential for data analysis in all meta-omics. The power of omics technologies depends on the capabilities of available bioinformatic tools in identifying and classifying microbial species, annotating sequences and proteins, predicting functional capabilities, and unveiling metabolic activities. Many bioinformatics algorithms and tools are available to analyze the omics data derived from various microbiomes. Most bioinformatics tools are initially developed for other microbiomes, and they are applied to omics investigations of the rumen microbiome. However, unlike other host-associated microbiomes, the rumen microbiome has diverse eukaryotic microbes (protozoa and fungi), which play significant roles in ruminant nutrition. These eukaryotic microbes are often overlooked and understudied due to the lack of appropriate bioinformatics tools. Bioinformatics algorithms employing machine learning are now available to analyze eukaryotes (Karlicki et al. 2022; Levy Karin et al. 2020; West et al. 2018). A recent bioinformatics tool GutEuk, developed specifically for rumen eukaryotes, can markedly enhance the analysis of these eukaryotes (Yan et al., 2024). Machine learning-based bioinformatics algorithms have also been developed to extract the largely underexplored viral sequence data (Du et al. 2023; Guo et al. 2021; Kieft et al. 2022; Nayfach et al. 2021) and mobile genetic elements (Tang et al. 2023). The advent of novel bioinformatics tools will greatly enhance comprehensive analyses of all domains and kingdoms within the rumen microbiome. It is envisaged that the near future will witness a substantial surge in data volumes capturing various facets of the rumen microbiome with unparalleled depth and resolution. Advancements in bioinformatics algorithms and tools, particularly those that can seamlessly integrate datasets from multi-omics sources, are needed to analyze this anticipated influx of diverse data effectively and adequately.

The experience from the preceding decades has shown that general-purpose genome databases have gaps, with inadequate representations of numerous microbial species. This deficiency becomes particularly evident when these databases are employed in rumen microbiome investigations. For example, a substantial portion of the biomass in the rumen is attributed to microbial eukaryotes, particularly protozoa (Andersen et al. 2023). However, the current databases have few genomes of rumen protozoa. Rumen viruses and fungi are also underrepresented in general-purpose databases. The recent bioinformatics tools tailored for viral sequence analysis have enabled the development of the first global comprehensive rumen virome database (Yan et al. 2023). A genome database of rumen protozoa must be developed for multi-omics investigations into this important group of rumen predators. The recently sequenced 52 single-cell amplified genomes (SAGs) are a valuable initial resource (Li et al. 2022b). Since zoospores of rumen fungi can be singularly picked, the single-cell genome sequencing approach used to sequence the SAGs of rumen protozoa may be used to sequence the genomes of rumen fungi.

Rumen microbiome nutriomics – connecting the rumen microbiome and nutrition

The intricate interplay among diet, the rumen microbiome, and ruminants establishes a dynamic nexus that forms the foundation for rumen functions, nutritional processes, and, ultimately, productivity. Investigating this nexus and identifying the rumen microbes or metabolic pathways that influence specific rumen functions or animal production traits has long been a focus of research. Through integrating multi-omics technologies and data analyses, rumen microbiome nutriomics can advance our comprehension of the roles played by rumen microbes in rumen functions and nutrition.

Rumen microbiome nutriomics through integrated omics and data analysis

Each omics technology has distinct capabilities and limitations. This recognition has led to the utilization of multiple omics in some recent studies, resulting in a more comprehensive characterization of the rumen microbiome (Liu et al. 2022a; Mu et al. 2022; Xu et al. 2021). However, few studies have sufficiently integrated the analysis of the data derived from different omics technologies or established the links between the data and rumen functions or animal production traits. This challenge is attributed, in part, to the multiple layers of high-dimensional microbiome data generated by individual omics technologies (Pedersen et al. 2018). Several strategies can reduce data dimensionality. These include combining data normalization, binning of co-abundant features (genes or metabolites), integration with prior biological knowledge (Pedersen et al. 2018), and clustering MAGs into metagenomic species (Zhang et al. 2023b). Additionally, identifying modules of related microbiome features, such as modules of microbiome, gene expression, and metabolites, can contribute to a more cohesive analysis. Bioinformatic approaches are continually evolving to integrate data derived from multi-omics technologies. For instance, a recent study utilized weighted gene co-expression network analysis and structural equation modeling (SEM) to integrate metataxonomic, metagenomic, and metabolomic data, revealing informative connections from rumen microbes to metabolites and milk protein yield (Zhang et al. 2023b). Furthermore, combinatorial network and machine learning methods have demonstrated utility in identifying metagenomic and host genotypes potentially linked to CH_4 emissions and feed efficiency in dairy cows (Cardinale and Kadarmideen 2022). In line with these advancements, we propose an integrated genome-centric and genome-resolved multi-omics approach to holistically characterize all the rumen microbes (i.e., prokaryotes, eukaryotes, and viruses) and key aspects of the rumen microbiome and establish connections with diets, rumen functions, and animal phenotype and production traits (Fig. 1).

In brief, existing high-quality MAGs, such as the large sets of MAGs of prokaryotes reported recently (Andersen et al. 2023; Stewart et al. 2019; Xie et al. 2021), viruses (Wu et al. 2024; Yan and Yu 2024), and genomes of the rumen microbiome such as those of the Hungate1000 project, ciliates (Li et al. 2022b; Park et al. 2021), and anaerobic fungi (Brown et al. 2021; Haitjema et al. 2017; Youssef et al. 2013), along with high-quality MAGs generated from ongoing studies, are combined to develop a comprehensive genome database (rumen microbiome genome database, RMGD). The MAGs and genomes of prokaryotes are taxonomically annotated using the taxonomy implemented in GTDB, which supports species-level classification based on the phylogeny derived from a concatenated set of 120 single-copy marker proteins. The RMGD is used for taxonomic classification and functional annotation of metagenomic and metatranscriptomic data. The RMGD can also be used in classifying operational taxonomic units (OTUs) or amplificon sequence variants (ASVs) generated

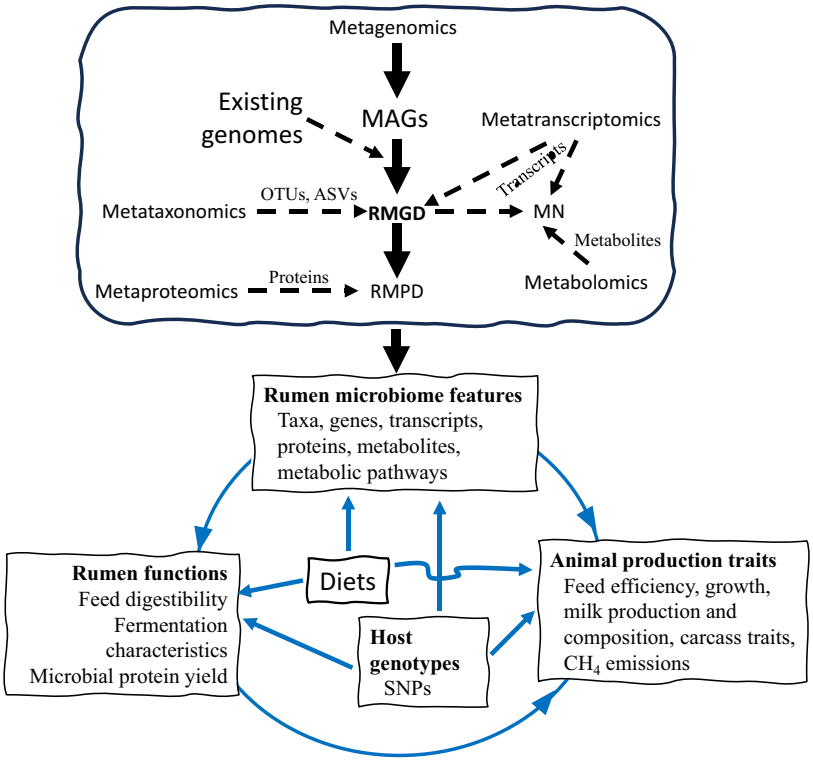

Figure 1. Integration of genome-centric and/or genome-resolved omics in investigating the nexus of rumen microbiome, functionalities, and animal production traits. A comprehensive rumen microbiome genome database (RMGD) is created using MAGS derived from genome-centric and genome-resolved metagenomics and genomes of rumen microbes. A rumen microbiome proteome database (RMPD) and metabolic networks (MN) are then prepared from the RMGD. Blue arrows indicate association analysis and causal inference. MAGs, metagenome-assembled genomes; OTUs, operational taxonomic units; ASVs, amplificon sequence variants; SNPs, single nucleotide polymorphisms.

by metataxonomics, potentially at the species level, by sequence mapping. The existing 52 SAGs of rumen ciliates (Li et al. 2022b) and the recent rumen virome database (RVD) (Yan et al. 2023) can be expanded to support rumen virome analysis. However, as discussed above, concerted efforts are needed to sequence more rumen protozoan and fungal genomes to develop a comprehensive rumen eukaryotic genome database and genome-based taxonomy.

The AA sequences translated from all the open reading frames (ORFs) of the RMGD are then used to prepare a rumen microbiome proteome database (RMPD) to aid in metaproteomic investigations of the rumen microbiome. Metabolic nextworks are assembled from the pathways reconstructed from individual MAGs and genomes to assist in identifying the transcripts and metabolites detected through metatranscriptomic and metabolomic analyses, respectively. The multiple layers of omics data are analyzed in an integrated manner (Subramanian et al. 2020). While having not been used in rumen microbiome nutriomics studies, xMWAS (Uppal et al. 2018) may be a valuable software for data integration, network visualization, clustering, and differential network analysis of data derived from two or more omics platforms. This integrated omics approach and data analysis will comprehensively characterize the rumen microbiome with respect to its many key features (Fig. 1). Furthermore, the integration of multiple omics data and analyses will enhance the accuracy of functional annotations. Such detailed data can be further interrogated in the context of diet, rumen functions, animal genotypes, and production traits.

Deciphering the interdependent labyrinth within the rumen ecosystem – Advancing toward establishing causality in the nutriomics of the rumen microbiome

The central goal of rumen microbiome nutriomics investigations is to delve into connections between various sets of data encompassing diets, rumen microbiome features identified through the omics technologies, rumen functions (i.e., feed digestion and fermentation characteristics), key production traits (e.g., feed digestibility, feed efficiency, growth, lactation performance, CH_4 emissions, etc.), and response to stress (e.g., heat stress), and nutritional disorders (e.g., subacute rumen acidosis). However, determining the causal relationships among these datasets remains an arduous task. Hence, researchers have used several analyses, such as DAA, correlation, and association analyses, to infer potential relationships. DAA can identify microbial taxa (primarily genera, OTUs, or ASV), functional categories of genes, and less frequently pathway enrichment, transcripts, and proteins that are differentially abundant between diets, animal groups, treatments, and animal production traits. Several analysis methods, including analysis of compositions of microbiomes with bias correction (ANCOM-BC), which address the data features of the rumen microbiome, in particular zero inflation and compositional effects, along with partial least squares discriminant analysis (PLS-DA) and linear discriminant analysis effect size (LEfSe), have been commonly used in DAA. Studies in ruminant nutrition frequently involve repeated measurements of the same subjects (for example, using a Latin square design) and experimental designs incorporating fixed and random effects (such as the randomized complete block design). For these studies, DAA methods capable of analyzing mixed effects, like LinDA (Zhou et al. 2022), should be used. However, all these methods have certain limitations (Nearing et al. 2022). To further improve DAA of microbiome data, some new methods that can better address the microbiome data features have been developed, such as ZicoSeq (Yang and Chen 2022), LOCOM (Hu et al. 2022), and CDEMI (Wang et al. 2023a). Future rumen microbiome nutriomics studies should employ these new methods. Like DAA, DGE analysis unveils variations in the expression of microbial genes and pathway enrichment; these variations can be associated with differences in diets, rumen

microbiome structure, rumen functions, or animal production traits. Correlations among these datasets can also be evaluated with appropriate, non-parametric methods. Differentially abundant microbial taxa and other microbiome features between, or those correlated with, specific rumen functions and animal production traits may guide further investigations into the causal relationships in rumen microbiome nutriomics.

Association analyses are used to reveal specific rumen microbiome taxa, often at the OTU (or ASV) and genus levels, as well as categories of functional genes linked to diets and animal production traits. Recently, microbiome, microbiota, or metagenome-wide association studies (MWAS) have been conducted to discover all taxa or functional gene categories associated with a specific host phenotype or disease status in humans and animals (Wang and Jia 2016). Despite having been frequently used to unveil gut microbes associated with diseases in humans (e.g., (Liu et al. 2021)) and feed efficiency (Aliakbari et al. 2022) along with intramuscular fat content (Wang et al. 2022b) in pigs, MWAS has only been used in a few recent studies on ruminants (Boggio et al. 2023a; Wang et al. 2023b). In a sheep study, MWAS did not identify any OTUs associated with dairy traits (Boggio et al. 2023a). The rumen microbiome has thousands of OTUs. Individual OTUs may lack sufficient "weight" to exhibit significant association. Therefore, MWAS might be more effectively applied to genera. Until now, MWAS has only been utilized to associate microbes detected through metataxonomics with animal production traits. However, assessing associations between rumen microbes and variations in diets and rumen functions will be equally applicable. Furthermore, MWAS should be able to examine associations between diet or animal production traits with rumen microbiome data derived from other omics technologies. An analogous approach, virome-wide association studies (VWAS), could be developed to identify rumen viruses associated with diets, specific rumen microbes, rumen functions, and animal production traits. This would represent a viral version of MWAS, extending the scope of broad association studies to include the viral component of the rumen ecosystem.

Many statistical or data analytics approaches, such as correlation, regression, probability, random forest, and deep learning, can be used in MWAS. However, the distinct data features of microbiomes may pose challenges to the robustness of MWAS. New methods are being developed to improve and streamline MWAS further. Recent examples include omnibus metagenome-wide association study with robustness (OMARU) (Kishikawa et al. 2022), MiATDS (Sun et al. 2021), and multiMiAT (Sun et al. 2023a). OMARU rigorously controls the statistical significance of MWAS results, including the correction of hidden confounding factors and the application of multiple test comparisons (Kishikawa et al. 2022). Additionally, OMARU can evaluate pathway-level links between metagenomes, as well as links between taxa and genes in metagenomes. MiATDS performs adaptive microbiome-based association analysis to detect microbial association signals with diverse sparsity levels (i.e., sparse, low sparse, non-sparse). This is achieved by defining the probability degree to measure the associations between microbes and host phenotypes and introducing the adaptive weighted sum of powered score tests by considering both probability degree and phylogenetic information (Sun et al. 2021). Divergently, implementing the multinomial logit model framework, multiMiAT supports MWAS between microbiomes and ordinal or nominal multicategory host phenotypes or traits (Sun et al. 2023a). Additionally, genome-wide association studies (GWAS) using rumen microbes as traits can identify heritable rumen microbes (Mani et al. 2022; Wang et al. 2023b). The integration of MWAS with GWAS, referred to as microbiome genome-wide association studies (mGWAS), provides a comprehensive approach for identifying heritable microbes associated with a specific phenotype (Wang et al. 2022a). For example, in growing lambs, mGWAS helped identify four genera of heritable rumen microbes associated with body weight (Wang et al. 2023b). This integrated methodology holds promise in identifying rumen microbes that can serve as potential markers, facilitating selective breeding for enhanced production traits.

It should be emphasized that association studies can only identify rumen microbes or other microbiome features exhibiting statistical correlations with diets, rumen functions, or specific animal production traits. Although prior knowledge of rumen microbes and their metabolism, meta-analyses of relevant studies, and longitudinal studies can help infer the biological plausibility of associations, caution must be exercised in interpreting these associations as causal relationships. To transcend mere correlation and association, it is crucial to establish causal relationships between diets, features of the rumen microbiome, rumen functions, and key animal production traits. Experimental testing or verification of the above causal relationships is arduous due to the complexity of the rumen microbiome and its intricate interactions with diet and animals. At the microbiome level, causality can be deduced through modeling approaches that integrate causality principles (Munoz-Tamayo et al. 2023). Several methods have been used to predict the microbes potentially driving a specific production trait in ruminants. These methodologies include SEM (Hertel et al. 2023) and causal Bayesian networks (CBNs) (Stebliankin et al. 2022). SEM allows researchers to investigate the direct and indirect effects of variables on one another, furnishing comprehensive insights into their complex relationships within a theoretical framework. On the other hand, CBNs utilize directed edges to represent causal relationships or data dependencies between variables, providing causal inference between rumen microbes and diet or animal production traits. SEM has proven valuable in identifying rumen microbial modules in dairy cows that potentially regulate milk protein yield (Zhang et al. 2023b) and CH_4 emissions (Saborio-Montero et al. 2020). While CBN analysis has not been used in rumen microbiome nutriomics studies, it has demonstrated utility in inferring causality between the gut microbiome and colorectal cancer (Kharrat et al. 2019) and between infant gut microbiome and resistome (Stebliankin et al. 2022) in humans. Causal inference models have also found applications in other human gut microbiome research (Hughes et al. 2020; Sun et al. 2023b). These analysis methods can be instrumental in deducing causal relationships in rumen microbiome nutriomics studies.

Machine learning has been increasingly used in investigating microbiomes, including rumen microbiomes. Compared with the traditional linear models commonly used in animal science research, machine learning is advantageous in analyzing large multidimensional data and inferring non-linear relationships. Machine learning has proven effective in predicting animal performance, exemplified by its application in forecasting CH_4 emissions from sheep (Zhang et al. 2023a), feed efficiency in dairy cows (Xue et al. 2022), and animal health conditions (Zhong et al. 2020) based on high-dimensional rumen microbiome data. Akin to its potential in human microbiome research, machine learning will prove itself to be useful in rumen microbiome nutriomics investigations, particularly in facilitating causal inference. Furthermore, recent studies have demonstrated the potential of artificial intelligence in human microbiome research (e.g., (Gao et al. 2023)). This suggests the prospect of applying artificial intelligence in

rumen microbiome nutriomics research, including discovering causality by analyzing large datasets of diets, rumen microbiomes, and animals and identifying their patterns and associations with animal production traits.

Assessing the magnitude of the rumen microbiome contribution to animal production traits

Because of its vital role in ruminant nutrition, the rumen microbiome significantly influences key animal production traits. Recent studies have quantitatively assessed the microbiability of specific animal traits, shedding light on the contribution of the rumen microbiome. Noteworthy among these traits are CH_4 emissions (m^2 of 13%) (Difford et al. 2018), RFI (m^2 of about 20%), milk energy (m^2 of about 25%) (Boggio et al. 2023b), and milk fatty acid composition in dairy cows ($m^2 > 30\%$ for some fatty acids) (Buitenhuis et al. 2019), as well as milk composition (m^2 of up to 7%) (Boggio et al. 2023a) and body weight (m^2 of 20%) in sheep (Wang et al. 2023b). These studies estimated the plausible contribution of the entire rumen microbiome to the production traits. Furthermore, MWAS based on single-OTU regression has revealed a small number of fecal OTUs significantly or suggestively linked to traits like RFI, feed conversion ratio (FCR), daily feed intake, and back fat thickness in pigs (Aliakbari et al. 2022). In a recent study utilizing machine learning to develop prediction models for CH_4 emissions from sheep, certain genera of rumen microbes were selected as predictor variables alongside animal data (Zhang et al. 2023a). The incorporation of microbial prediction variables not only enhanced prediction accuracy but also bolstered model robustness. This machine learning approach identifies rumen microbes potentially associated with CH_4 emissions and provides insights into their effect sizes through the coefficients of the microbial predictor variables.

No study has assessed the microbiability of rumen functions, such as feed digestibility and VFA profiles. Significant microbiability has been demonstrated for digestive efficiency in pigs (Deru et al. 2022). Given the direct correlation between rumen functions, feed efficiency, and other key animal production traits, future research is warranted to investigate the microbiability of major rumen functions and the rumen microbial taxa (genera or species) contributing to those functions. Furthermore, heritable rumen bacteria contribute more to the microbiability of lactation performance than their nonheritable counterparts (Zang et al. 2022). Given that host genetics significantly influence heritable rumen microbes, a novel metric called "holobiality" (ho^2) has been proposed. This metric combines the heritability of specific production traits with their microbial contribution (microbiability). By quantifying the joint influence of the host genome and rumen microbiome, holobiality offers promising potential for predicting improvement in these traits.

Concluding remarks and future perspectives

The ruminant industry faces challenges in optimizing feed efficiency, minimizing environmental impact, and enhancing the quality of milk and meat to meet growing global demands for dairy and meat products. The diverse rumen microbiome, as the primary supplier of metabolizable energy, protein, and precursors of milk and muscle protein, requires a more profound understanding with respect to its composition and functions, as well as its interactions with diet, animal genotypes, and production traits. Furthermore, it is essential to unveil the causal relationship among these layers of variables, including data from ruminants. While various omics technologies have been used to investigate the rumen microbiome, each has limitations. Recently, the use of two or three multi-omics has shown promise in providing a more comprehensive characterization of the rumen microbiome. However, the data generated through these technologies are often not sufficiently analyzed or interpreted in an integrated manner. The proposed concept of rumen microbiome nutriomics advocates for the integration of multiple omics and data analyses. This integrated approach aims to advance our comprehensive understanding of the rumen microbiome and its intricate interactions with both diet and animals. The establishment of the Animal Nutriomics journal serves as a vital platform for disseminating and exchanging novel findings from rumen microbiome nutriomics investigations, facilitating collaboration and knowledge dissemination within the scientific community.

Central to rumen microbiome nutriomics is the development of a comprehensive RMGD. While some researchers have developed in-house databases, there exists significant variability in their completeness and accuracy of curation, thereby exerting a significant influence on the analysis outcomes (Smith et al. 2022). A serious undertaking is needed to compile the genomes and high-quality MAGs of rumen microbes into a publicly accessible RMGD. Concerted efforts among researchers in the realm of the rumen microbiome are needed to sequence more genomes of rumen protozoa, fungi (in particular), and viruses or mine the existing rumen metagenomes. Moreover, comparisons of results and findings among studies have been compromised or difficult due to the lack of sufficient metadata and technical variations across studies, such as study design; sampling; extraction or isolation of DNA, RNA, protein, and metabolites; as well as bioinformatic analyses (Hagey et al. 2022; McGovern et al. 2018; Rintala et al. 2017). A set of criteria for the above technical aspects and workflows will be valuable to standardize the processes of rumen microbiome nutriomics investigations. Such standardization will be particularly invaluable for data reanalysis using big data analytics.

The "holy grail" of rumen microbiome nutriomics is to reveal and understand the causal relationships between different layers of data, ranging from diet, rumen microbiome, rumen functions, animal genotypes, and key production traits. DAA, correlation, and association analysis may help identify potential interactions and indicators of some important aspects, such as feed efficiency, product quality, or methane emissions. However, causal inference is urgently needed to establish their cause–effect relationships. Several approaches, including modeling and CBNs, can be used. Machine learning and artificial intelligence also hold potential in this pursuit in future investigations.

References

Abellan-Schneyder I, Siebert A, Hofmann K *et al.* (2021) Full-length SSU rRNA gene sequencing allows species-level detection of bacteria, archaea, and yeasts present in milk. *Microorganisms* **9**(6), 1251.

Ali S, X. N, Khan M *et al.* (2023) Effects of dietary protein levels on sheep gut metabolite profiles during the lactating stage. *Animals (Basel)* **14**(1), 121.

Aliakbari A, Zemb O, Cauquil L *et al.* (2022) Microbiability and microbiome-wide association analyses of feed efficiency and performance traits in pigs. *Genetics Selection Evolution* **54**(1), 29.

Amin AB, Zhang L, Zhang J *et al.* (2022) Metagenomic and metabolomic insights into the mechanism underlying the disparity in milk yield of Holstein cows. *Frontiers in Microbiology* **13**, 844968.

Andersen TO, Altshuler I, Vera-ponce de Leon A *et al.* (2023) Metabolic influence of core ciliates within the rumen microbiome. *ISME J* **17**(7), 1128–1140.

Andersen TO, Kunath BJ, Hagen LH *et al.* (2021) Rumen metaproteomics: Closer to linking rumen microbial function to animal productivity traits. *Methods* **186**, 42–51.

Anderson CL, Sullivan MB and Fernando SC (2017) Dietary energy drives the dynamic response of bovine rumen viral communities. *Microbiome* **5**(1), 155.

Bhagat NR, Kumar S, Kumari R *et al.* (2023) A review on rumen anaerobic fungi: Current understanding on carbohydrate fermentation and roughages digestion in ruminants. *Applied Biochemistry and Microbiology* **59**(3), 231–249.

Bileschi ML, Belanger D, Bryant DH *et al.* (2022) Using deep learning to annotate the protein universe. *Nature Biotechnology* **40**(6), 932–937.

Boggio GM, Christensen OF, Legarra A *et al.* (2023a) Microbiability of milk composition and genetic control of microbiota effects in sheep. *Journal of Dairy Science* **106**(9), 6288–6298.

Boggio GM, Monteiro HF, Lima FS *et al.* (2023b) Host and rumen microbiome contributions to feed efficiency traits in Holstein cows. *Journal of Dairy Science* **107**(5), 3090–3103.

Brown JL, Swift CL, Mondo SJ *et al.* (2021) Co-cultivation of the anaerobic fungus *Caecomyces churrovis* with *Methanobacterium bryantii* enhances transcription of carbohydrate binding modules, dockerins, and pyruvate formate lyases on specific substrates. *Biotechnology for Biofuels* **14**(1), 234.

Brulc JM, Yeoman CJ, Wilson MK *et al.* (2011) Cellulosomics, a gene-centric approach to investigating the intraspecific diversity and adaptation of *Ruminococcus flavefaciens* within the rumen. *PLoS ONE* **6**(10), e25329.

Buchfink B, Xie C and Huson DH (2015) Fast and sensitive protein alignment using diamond. *Nature Methods* **12**(1), 59–60.

Buitenhuis B, Lassen J, Noel SJ *et al.* (2019) Impact of the rumen microbiome on milk fatty acid composition of Holstein cattle. *Genetics Selection Evolution* **51**(1), 23.

Callahan BJ, Grinevich D, Thakur S *et al.* (2021) Ultra-accurate microbial amplicon sequencing with synthetic long reads. *Microbiome* **9**(1), 130.

Cardinale S and Kadarmideen HN (2022) Host genome-metagenome analyses using combinatorial network methods reveal key metagenomic and host genetic features for methane emission and feed efficiency in cattle. *Frontiers in Genetics* **13**, 795717.

Chaumeil PA, Mussig AJ, Hugenholtz P *et al.* (2022) GTDB-tk v2: Memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**(23), 5315–5316.

Chklovski A, Parks DH, Woodcroft BJ *et al.* (2022) CheckM2: A rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. bioRxiv:2022.07.11.499243.

Creevey CJ, Kelly WJ, Henderson G *et al.* (2014) Determining the culturability of the rumen bacterial microbiome. *Microbial Biotechnology* **7**(5), 467–479.

Delgado LF and Andersson AF (2022) Evaluating metagenomic assembly approaches for biome-specific gene catalogues. *Microbiome* **10**(1), 72.

Denman SE, Morgavi DP and McSweeney CS (2018) Review: The application of omics to rumen microbiota function. *Animal* **12**(s2), s233–s45.

de Poppi AC, Lazzari G, Gomes ALM *et al.* (2021) Effects of feeding a live yeast on rumen fermentation and fiber degradability of tropical and subtropical forages. *Journal of the Science of Food & Agriculture* **101**(15), 6220–6227.

Deru V, Tiezzi F, Carillier-Jacquin C *et al.* (2022) Gut microbiota and host genetics contribute to the phenotypic variation of digestive and feed efficiency traits in growing pigs fed a conventional and a high fiber diet. *Genetics Selection Evolution* **54**(1), 55.

Difford GF, Plichta DR, Lovendahl P *et al.* (2018) Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genetetics* **14**(10), e1007580.

Douglas GM, Maffei VJ, Zaneveld JR *et al.* (2020) PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology* **38**(6), 685–688.

Du Y, Fuhrman JA and Sun F (2023) ViralCC retrieves complete viral genomes and virus-host pairs from metagenomic Hi-C data. *Nature Communications* **14**(1), 502.

Elekwachi CO, Wang Z, X. W *et al.* (2017) Total rRNA-seq analysis gives insight into bacterial, fungal, protozoal and archaeal communities in the rumen using an optimized RNA isolation method. *Frontiers in Microbiology* **8**, 1814.

Eom JS, Kim HS, Lee SJ *et al.* (2021) Metabolic profiling of rumen fluid and milk in lactating dairy cattle influenced by subclinical ketosis using proton nuclear magnetic resonance spectroscopy. *Animals (Basel)* **11**(9), 2526.

Feng L, Zhang Y, Liu W *et al.* (2022) Effects of heat stress on 16S rDNA, metagenome and metabolome in Holstein cows at different growth stages. *Scientific Data* **9**(1), 644.

Frioux C, Singh D, Korcsmaros T *et al.* (2020) From bag-of-genes to bag-of-genomes: Metabolic modelling of communities in the era of metagenome-assembled genomes. *Computational and Structural Biotechnology Journal* **18**, 1722–1734.

Gao S, Gao X, Zhu R *et al.* (2023) Microbial genes outperform species and SNVs as diagnostic markers for Crohn's disease on multicohort fecal metagenomes empowered by artificial intelligence. *Gut Microbes* **15**(1), 2221428.

Gilbert RA, Townsend EM, Crew KS *et al.* (2020) Rumen virus populations: Technological advances enhancing current understanding. *Frontiers in Microbiology* **11**, 450.

Guo J, Bolduc B, Zayed AA *et al.* (2021) Virsorter2: A multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**(1), 37.

Hagey JV, Laabs M, Maga EA *et al.* (2022) Rumen sampling methods bias bacterial communities observed. *PLoS ONE* **17**(5), e0258176.

Haitjema CH, Gilmore SP, Henske JK *et al.* (2017) A parts list for fungal cellulosomes revealed by comparative genomics. *Nature Microbiology* **2**(8), 17087.

Hertel J, Heinken A, Fassler D *et al.* (2023) Causal inference on microbiome-metabolome relations in observational host-microbiome data via in silico in vivo association pattern analyses. *Cell Reports Methods* **3**(10), 100615.

Heyer R, Schallert K, Zoun R *et al.* (2017) Challenges and perspectives of metaproteomic data analysis. *Journal of Biotechnology* **261**, 24–36.

Hu Y, Satten GA and Hu YJ (2022) LOCOM: A logistic regression model for testing differential abundance in compositional microbiome data with false discovery rate control. *Proceedings of the National Academy of Sciences of the United States of America* **119**(30), e2122788119.

Hughes DA, Bacigalupe R, Wang J *et al.* (2020) Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. *Nature Microbiology* **5**(9), 1079–1087.

Idowu M, Taiwo G, Sidney T *et al.* (2023) The differential plasma and ruminal metabolic pathways and ruminal bacterial taxa associated with divergent residual body weight gain phenotype in crossbred beef steers. *Translational Animal Science* **7**(1), txad054.

Jize Z, Zhuoga D, Xiaoqing Z *et al.* (2022) Different feeding strategies can affect growth performance and rumen functions in Gangba sheep as revealed by integrated transcriptome and microbiome analyses. *Frontiers in Microbiology* **13**, 908326.

Johnson JS, Spakowicz DJ, Hong BY *et al.* (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications* **10**(1), 5029.

Joseph TA, Chlenski P, Litman A *et al.* (2022) Accurate and robust inference of microbial growth dynamics from metagenomic sequencing reveals personalized growth rates. *Genome Research* **32**(3), 558–568.

Karlicki M, Antonowicz S and Karnkowska A (2022) Tiara: Deep learning-based classification system for eukaryotic sequences. *Bioinformatics* **38**(2), 344–350.

Kharrat N, Assidi M, Abu-Elmagd M *et al.* (2019) Data mining analysis of human gut microbiota links *Fusobacterium* spp. with colorectal cancer onset. *Bioinformation* **15**(6), 372–379.

Kieft K, Adams A, Salamzade R *et al.* (2022) vRhyme enables binning of viral genomes from metagenomes. *Nucleic Acids Research* **50**(14), e83.

Kieft K, Zhou Z and Anantharaman K (2020) Vibrant: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**(1), 90.

Kim M, Morrison M and Yu Z (2011) Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiology Ecology* **76**(1), 49–63.

Kishikawa T, Tomofuji Y, Inohara H *et al.* (2022) OMARU: A robust and multifaceted pipeline for metagenome-wide association study. *NAR Genom Bioinform* **4**(1), lqac019.

Korem T, Zeevi D, Suez J *et al.* (2015) Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**(6252), 1101–1106.

Laursen MF, Dalgaard MD and Bahl MI (2017) Genomic GC-content affects the accuracy of 16S rRNA gene sequencing based microbial profiling due to PCR bias. *Frontiers in Microbiology* **8**, 1934.

Levy Karin E, Mirdita M and Soding J (2020) MetaEuk-sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome* **8**(1), 48.

Li F, Hitch TCA, Chen Y *et al.* (2019) Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the rumen microbiome and its associations with feed efficiency in beef cattle. *Microbiome* **7**(1), 6.

Li S, Zeng H, Wang C *et al.* (2022a) Effect of methionine hydroxy analog feed supplements: Significant alteration and enrichment of rumen microbiota and metabolome in Hu sheep. *Frontiers in Veterinary Science* **9**, 999726.

Li Y, Mao K, Zang Y *et al.* (2023) Revealing the developmental characterization of rumen microbiome and its host in newly received cattle during receiving period contributes to formulating precise nutritional strategies. *Microbiome* **11**(1), 238.

Li Y, Zang Y, Zhao X *et al.* (2021) Dietary supplementation with creatine pyruvate alters rumen microbiota protein function in heat-stressed beef cattle. *Frontiers in Microbiology* **12**, 715088.

Li Z, Wang X, Zhang Y *et al.* (2022b) Genomic insights into the phylogeny and biomass-degrading enzymes of rumen ciliates. *The ISME Journal* **16**(12), 2775–2787.

Liu P, Jiang Y, Gu S *et al.* (2021) Metagenome-wide association study of gut microbiome revealed potential microbial marker set for diagnosis of pediatric *Myasthenia gravis*. *BMC Medicine* **19**(1), 159.

Liu X, Sha Y, W. L *et al.* (2022a) Multi-omics reveals that the rumen transcriptome, microbiome, and its metabolome co-regulate cold season adaptability of Tibetan sheep. *Frontiers in Microbiology* **13**, 859601.

Liu Y, Wu H, Chen W *et al.* (2022b) Rumen microbiome and metabolome of high and low residual feed intake Angus heifers. *Frontiers in Veterinary Science* **9**, 812861.

Lohmann P, Schape SS, Haange SB *et al.* (2020) Function is what counts: How microbial community complexity affects species, proteome and pathway coverage in metaproteomics. *Expert Review of Proteomics* **17**(2), 163–173.

Lu J and Salzberg SL (2020) Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome* **8**(1), 124.

Mani S, Aiyegoro OA and Adeleke MA (2022) Association between host genetics of sheep and the rumen microbial composition. *Tropical Animal Health and Production* **54**(2), 109.

Martinez Boggio G, Meynadier A, Buitenhuis AJ *et al.* (2022) Host genetic control on rumen microbiota and its impact on dairy traits in sheep. *Genetic Selection Evolution* **54**(1), 77.

McDonald D, Jiang Y, Balaban M *et al.* (2023) Greengenes2 unifies microbial data in a single reference tree. *Nature Biotechnology*. Online ahead of print. doi:10.1038/s41587-023-02026-w

McGovern E, Waters SM, Blackshields G *et al.* (2018) Evaluating established methods for rumen 16S rRNA amplicon sequencing with mock microbial populations. *Frontiers in Microbiology* **9**, 1365.

Menzel P, Ng KL and Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* **7**(1), 11257.

Milanese A, Mende DR, Paoli L *et al.* (2019) Microbial abundance, activity and population genomic profiling with mOTUs2. *Nature Communications* **10**(1), 1014.

Mirdita M, Steinegger M, Breitwieser F *et al.* (2021) Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics* **37**(18), 3029–3031.

Miura N and Okuda S (2023) Current progress and critical challenges to overcome in the bioinformatics of mass spectrometry-based metaproteomics. *Computational and Structural Biotechnology Journal* **21**, 1140–1150.

Mu Y, Qi W, Zhang T *et al.* (2022) Multi-omics analysis revealed coordinated responses of rumen microbiome and epithelium to high-grain-induced subacute rumen acidosis in lactating dairy cows. *mSystems* **7**(1), e0149021.

Munoz-Tamayo R, Davoudkhani M, Fakih I *et al.* (2023) Review: Towards the next-generation models of the rumen microbiome for enhancing predictive power and guiding sustainable production strategies. *Animal* **17 Suppl 5**, 100984.

Nayfach S, Camargo AP, Schulz F *et al.* (2021) CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology* **39**(5), 578–585.

Nearing JT, Douglas GM, Hayes MG *et al.* (2022) Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications* **13**(1), 342.

Nelson WC, Tully BJ and Mobberley JM (2020) Biases in genome reconstruction from metagenomic data. *PeerJ* **8**, e10119.

Olm MR, Crits-Christoph A, Bouma-Gregson K *et al.* (2021) Instrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nature Biotechnology* **39**(6), 727–736.

Park T, Cersosimo LM, Radloff W *et al.* (2022) The rumen liquid metatranscriptome of post-weaned dairy calves differed by pre-weaning ruminal administration of differentially-enriched, rumen-derived inocula. *Anim Microbiome* **4**(1), 4.

Park T, Wijeratne S, Meulia T *et al.* (2021) The macronuclear genome of anaerobic ciliate *Entodinium caudatum* reveals its biological features adapted to the distinct rumen environment. *Genomics* **113**(3), 1416–1427.

Parks DH, Chuvochina M, Waite DW *et al.* (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* **36**(10), 996–1004.

Pedersen HK, Forslund SK, Gudmundsdottir V *et al.* (2018) A computational framework to integrate high-throughput '-omics' datasets for the identification of potential mechanistic links. *Nature Protocols* **13**(12), 2781–2800.

Pitta DW, Indugu N, Melgar A *et al.* (2022) The effect of 3-nitrooxypropanol, a potent methane inhibitor, on ruminal microbial gene expression profiles in dairy cows. *Microbiome* **10**(1), 146.

Ren Y, Zhaxi Y, Ciwang R *et al.* (2023) Responses of rumen microorganisms and metabolites to different roughage of domesticated Tibetan sheep. *Frontiers in Microbiology* **14**, 1247609.

Rezaeian M, Beakes GW and Parker DS (2004) Distribution and estimation of anaerobic zoosporic fungi along the digestive tracts of sheep. *Mycological Research* **108**(Pt 10), 1227–1233.

Rintala A, Pietila S, Munukka E *et al.* (2017) Gut microbiota analysis results are highly dependent on the 16S rRNA gene target region, whereas the impact of DNA extraction is minor. *Journal of Biomolecular Techniques* **28**(1), 19–30.

Saborio-Montero A, Gutierrez-Rivas M, Garcia-Rodriguez A *et al.* (2020) Structural equation models to disentangle the biological relationship between microbiota and complex traits: Methane production in dairy cattle as a case of study. *Journal of Animal Breeding and Genetics* **137**(1), 36–48.

Shakya M, Lo CC and Chain PSG (2019) Advances and challenges in metatranscriptomic analysis. *Frontiers in Genetics* **10**, 904.

Silverman JD, Bloom RJ, Jiang S *et al.* (2021) Measuring and mitigating PCR bias in microbiota datasets. *PLoS Computational Biology* **17**(7), e1009113.

Smith RH, Glendinning L, Walker AW *et al.* (2022) Investigating the impact of database choice on the accuracy of metagenomic read classification for the rumen microbiome. *Animal Microbiome* **4**(1), 57.

Somasundaram S and Yu Z (2024) Internal transcribed spacers as phylogenetic markers enable species-level metataxonomic analysis of ciliated protozoa. bioRxiv:2024.02.02.578691.

Stebliankin V, Sazal M, Valdes C *et al.* (2022) A novel approach for combining the metagenome, metaresistome, metareplicome and causal inference to determine the microbes and their antibiotic resistance gene repertoire that contribute to dysbiosis. *Microbial Genomics* **8**(12), mgen000899.

Stewart RD, Auffret MD, Warr A *et al.* (2019) Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nature Biotechnology* **37**(8), 953–961.

Subramanian I, Verma S, Kumar S *et al.* (2020) Multi-omics data integration, interpretation, and its application. *Bioinformatics and Biology Insights* **14**, 1177932219899051.

Sun H, Huang X, Fu L *et al.* (2021) A powerful adaptive microbiome-based association test for microbial association signals with diverse sparsity levels. *Journal of Genetics and Genomics* **48**(9), 851–859.

Sun H, Wang Y, Xiao Z *et al.* (2023a) MultiMiAT: An optimal microbiome-based association test for multicategory phenotypes. *Briefings in Bioinformatics* **24**(2), bbad012.

Sun K, Gao Y, Wu H *et al.* (2023b) The causal relationship between gut microbiota and type 2 diabetes: A two-sample Mendelian randomized study. *Frontiers in Public Health* **11**, 1255059.

Tang X, Shang J, Ji Y *et al.* (2023) PLASMe: A tool to identify plasmid contigs from short-read assemblies using transformer. *Nucleic Acids Research* **51**(15), e83.

Trautmann A, Schleicher L, Koch A *et al.* (2023) A shift towards succinate-producing *Prevotella* in the ruminal microbiome challenged with monensin. *Proteomics* **23**(21-22), e2200121.

Tremblay J, Singh K, Fern A *et al.* (2015) Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology* **6**, 771.

Truong DT, Franzosa EA, Tickle TL *et al.* (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods* **12**(10), 902–903.

Uppal K, Ma C, Go YM *et al.* (2018) xMWAS: A data-driven integration and differential network analysis tool. *Bioinformatics* **34**(4), 701–702.

von Meijenfeldt FAB, Arkhipova K, Cambuy DD *et al.* (2019) Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology* **20**(1), 217.

Wang J and Jia H (2016) Metagenome-wide association studies: Fine-mining the microbiome. *Nature Reviews, Microbiology* **14**(8), 508–522.

Wang L, Liang X, Chen H *et al.* (2023a) CDEMI: Characterizing differences in microbial composition and function in microbiome data. *Computational and Structural Biotechnology Journal* **21**, 2502–2513.

Wang W, Zhang Y, Zhang X *et al.* (2023b) Heritability and recursive influence of host genetics on the rumen microbiota drive body weight variance in male Hu sheep lambs. *Microbiome* **11**(1), 197.

Wang Y, Wang X, Sun S *et al.* (2022a) GWAS, MWAS and mGWAS provide insights into precision agriculture based on genotype-dependent microbial effects in foxtail millet. *Nature Communications* **13**(1), 5913.

Wang Y, Zhou P, Zhou X *et al.* (2022b) Effect of host genetics and gut microbiome on fat deposition traits in pigs. *Frontiers in Microbiology* **13**, 925200.

West PT, Probst AJ, Grigoriev IV *et al.* (2018) Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Research* **28**(4), 569–580.

Wilkinson TJ, Huws SA, Edwards JE *et al.* (2018) CowPI: A rumen microbiome focussed version of the PICRUSt functional inference software. *Frontiers in Microbiology* **9**, 1095.

Wu Y, Gao N, Sun C *et al.* (2024) A compendium of ruminant gastrointestinal phage genomes revealed a higher proportion of lytic phages than in any other environments. *Microbiome* **12**(1), 69.

Xie F, Jin W, H. S *et al.* (2021) An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome* **9**(1), 137.

Xu C, Liu W, Sun B *et al.* (2021) Multi-omics analysis reveals a dependent relationship between rumen bacteria and diet of grass-and grain-fed yaks. *Frontiers in Microbiology* **12**, 642959.

Xue MY, Xie YY, Zhong Y *et al.* (2022) Integrated meta-omics reveals new ruminal microbial features associated with feed efficiency in dairy cattle. *Microbiome* **10**(1), 32.

Yan M, Andersen T, Pope P *et al.* (2024). Probing the eukaryotic microbes of ruminants with a deep-learning classifier and comprehensive protein databases. *bioRxiv*. doi:10.1101/2024.07.17.603995

Yan M, Pratama AA, Somasundaram S *et al.* (2023) Interrogating the viral dark matter of the rumen ecosystem with a global virome database. *Nature Communications* **14**(1), 5254.

Yan M and Yu Z (2024) Viruses contribute to microbial diversification in the rumen ecosystem and are associated with certain animal production traits. *Microbiome*. **12**(1), 82.

Yang L and Chen J (2022) A comprehensive evaluation of microbial differential abundance analysis methods: Current status and potential solutions. *Microbiome* **10**(1), 130.

Youssef NH, Couger MB, Struchtemeyer CG *et al.* (2013) The genome of the anaerobic fungus *Orpinomyces* sp. strain C1A reveals the unique evolutionary history of a remarkable plant biomass degrader. *Applied and Environmental Microbiology* **79**(15), 4620–4634.

Yu Z and Morrison M (2004) Comparisons of different hypervariable regions of *rrs* genes for use in fingerprinting of microbial communities by PCR-denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology* **70**(8), 4800–4806.

Zang XW, Sun HZ, Xue MY *et al.* (2022) Heritable and nonheritable rumen bacteria are associated with different characters of lactation performance of dairy cows. *mSystems* **7**(5), e0042222.

Zhang B, Lin S, Moraes L *et al.* (2023a) Methane prediction equations including genera of rumen bacteria as predictor variables improve prediction accuracy. *Scientiifc Reports* **13**(1), 21305.

Zhang B, Lin S and Yu Z (2022) Metagenome-predicted growth rate and metatranscriptomic analysis reveal a slow growth rate of and high butyrate formation by *Faecalibacterium prausnitzii* in the rumen of low methane-emitting sheep. *Research Square*. Preprint. doi:10.21203/rs.3.rs-1450754/v1

Zhang C, Wang M, Liu H *et al.* (2023b) Multi-omics reveals that the host-microbiome metabolism crosstalk of differential rumen bacterial enterotypes can regulate the milk protein synthesis of dairy cows. *Journal of Animal Science and Biotechnology* **14**(1), 63.

Zhang Y, F. L, Chen Y *et al.* (2020) Metatranscriptomic profiling reveals the effect of breed on active rumen eukaryotic composition in beef cattle with varied feed efficiency. *Frontiers in Microbiology* **11**, 367.

Zhong Y, Xue MY, Sun HZ *et al.* (2020) Rumen and hindgut bacteria are potential indicators for mastitis of mid-lactating Holstein dairy cows. *Microorganisms* **8**(12), 2042.

Zhou H, K. H, Chen J *et al.* (2022) Linda: Linear models for differential abundance analysis of microbiome compositional data. *Genome Biology* **23**(1), 95.