EFFECT OF EARLY-LIFE MANAGEMENT PRACTICES ON THE RUMEN METAGENOME OF BEEF CATTLE

A THESIS
SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA
BY

Gerardo R Diaz Ortiz

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

Advised by: Noelle R Noyes, DVM, PhD, MA

2023

ACKNOWLEDGEMENTS

This research was funded by the Minnesota Beef Council using Beef Checkoff funds. My Master's studies were partially funded by Fulbright (Grantee No. GNT-20212290), *Consejo Nacional de Ciencia, Tecnologia e Innovacion Tecnologica* from Peru (Contract No. 085-2020-FONDECYT) and a MnDRIVE Graduate Student Professional Development award (MnDRIVE Global Food Ventures Program).

I would like to thank Tui Ray, Mojtaba Daneshi and Ilya Slizovskiy, for their valuable help during sample collection and DNA extraction, and to Peter Ferm and Tara Gaire for their patience and guidance in the bioinformatics and statistical analysis of this work. I acknowledge my Master thesis committee members, Dr. Andres Gomez and Dr. Timothy Goldsmith for their insightful comments and feedback that shaped this scientific work. My deepest sense of gratitude to my advisor, Noelle Noyes, for her unconditional support, warm personality and inspirational leadership that showed me another way to do science: striving for excellence with transparency, partnership, and enthusiasm. Special acknowledgements to the University of Minnesota Genomics Center (UMGC), the Minnesota Supercomputing Institute (MSI) and the North Central Research and Outreach Center (NCROC) at the University of Minnesota for providing the necessary support and resources to carry out this research.

Even more important than science are friends and family. I am grateful to have met fantastic friends during this journey: My MGC friends, my lab friends, and my Latino community at CTC. I would like to thank them for their friendship and great moments that cheered me up when I felt homesick. My final and utmost expression of gratitude is reserved for my family, both in Peru and here, to whom I owe everything that I am. Thanks to my 2 mothers, Nelly and Feli, my uncles, aunts, and siblings, Ivan, Buka and Carmen, who raised me as a nonconformist man, always looking for new horizons. Thanks to my wife, Marilia, and my daughter, Jimena, who believed in me even more than I do and have generously shared a part of their own lives, as proof of their purest love and trust.

DEDICATION

Para Marilia y Jimena,
mis eternas compañeras.

ABSTRACT

Although the rumen microbiome plays a critical role in beef cattle health and productivity, the effect of early-life management practices on the rumen metagenome of beef calves has been scarcely studied. Through two research studies, this thesis investigated the effect of four castration time windows and two weaning strategies on the rumen metagenome of beef calves through a comprehensive metagenomic sequencing approach that included a technical comparison of shotgun metagenomic sequencing and 16S rRNA gene sequencing to study the rumen microbiome. We found that castration timing had limited long-term effects on the rumen microbiome, while weaning strategy showed short-term effects on the microbiome composition and methanogenic gene content of the rumen, but not on the rumen resistome. In addition, our results showed that despite technical discrepancies, 16S rRNA gene sequencing and shotgun metagenomic sequencing methods led to similar ecological inferences about the effect of weaning strategy on the rumen microbiome of beef calves. These studies highlight the importance of considering early-life interventions in beef cattle management and provide insights into the comparability of rumen microbiome sequencing methods. Understanding rumen microbiome dynamics and choosing appropriate sequencing approaches are crucial for advancing rumen microbiome research in beef cattle.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 01: LITERATURE REVIEW

Within the production cycle of beef cattle, the cow-calf operation is a critical phase as it represents most of the production costs (~60%) [1], has high associated methane emissions (~77%) [2] and has the highest mortality rate across the production cycle (~6.4%) [3]. The cow-calf operation is characterized by an extensive pasture-based feeding system that produces weaned calves to stockers, backgrounding or feedlot systems [4]. Several management practices are implemented to optimize the production and health of the animals in the cow-calf stage, some of them having a long-term impact on animal productivity [5]. Due to its central role in beef cattle physiology, the microbial community of the rumen has been deeply investigated in the last decades using culture-independent sequencing methods [6]. 16S rRNA gene sequencing, and recently shotgun metagenomic sequencing, have allowed a comprehensive description of the taxonomy, diversity, and putative function of the rumen microbial community, also known as the rumen microbiome.

Unlike dairy cattle, the early-life dynamics of the beef cattle rumen microbiome and the impact of management practices on it are scarcely documented. Furthermore, the various culture-independent approaches used to describe the rumen microbiome have not been robustly compared, which leads to a lack of understanding about whether or how we can compare results across different studies that use different approaches. To address these knowledge gaps, this review will focus on (1) current knowledge about the rumen microbiome of beef cattle and its association with production traits, (2) a description of the sequencing methods used to study the rumen microbiome of beef cattle, and (3) the impact of management practices on the rumen microbiome, primarily focused on early life (cow-calf stage) of beef cattle.

ASSOCIATION OF THE RUMEN MICROBIOME WITH PRODUCTION TRAITS OF BEEF CATTLE

General description of the cattle rumen microbiome

The cattle rumen microbiome is a specialized microbial community made up of bacteria, fungi, archaea, protozoa and viruses with cellulolytic, hemicellulolytic, amylolytic,

proteolytic, and lipolytic functions [7]. These microorganisms transform fibrous feedstuff into volatile fatty acids (VFAs) such as acetate, propionate, and butyrate [8] in a 3-level cascade process that includes: (1) cellulose and hemicellulose degradation into oligosaccharides via glycosyl hydrolase enzymes, (2) soluble sugar fermentation mainly via Embden–Meyerhoff–Parnas pathway, pentose phosphate pathway and anaerobic fermentation, and (3) use of fermentation products, specifically organic acids to produce VFAs, and hydrogen and methanol to produce methane and acetate via hydrogenotrophic, methylotrophic and acetoclastic pathways [6]. VFAs provide the host with around 70% of its metabolizable energy to produce human edible products, such as meat and milk [6,8].

Bacteria are the most abundant members of the rumen microbiome with more than 50% of the microbial biomass, whereas archaea account for less than 4% of this biomass [8]. A multisite study in Europe found that a small proportion of microbial species regardless of geographic location and diet can predict animal productivity, rumen metabolites and methane emissions in dairy cows. These few but highly abundant species (0.25% of the overall species with 30-60% overall abundance) included the cellulose degraders *Fibrobacter succinogenes, Ruminococcus flavefaciens* and *Ruminococcus albus*, and hemicellulose degraders within the *Prevotella, Butyrivibrio* and *Pseudobutyrivibrio* genera [9]. Other important bacterial species included *Treponema bryantii* and *Selenomonas ruminantium* which use fermentation products [6], and *Sphaerochaeta* spp. which typically produces acetate [10]. The most important archaeal genera were *Methanobrevibacter* spp., a methanogenic taxon that accounts for up to 70% of the ruminal archaeal community. *Methanobrevibacter* spp. are known as hydrogen (electron) sinks because they transform free hydrogen into methane, which benefits the hydrolytic activity of cellulose degraders, inhibited sometimes by saturated hydrogen concentration in the rumen [6] **(Table 1.1).**

Ciliate protozoa and fungi are other important members of the rumen microbiome that account for up to 50% and between 10-20% of the microbial biomass, respectively [11]. The fungal genera *Lewia, Neocallimastix*, *Phoma*, *Alternaria, Candida*, and *Piromyces* spp. [12], and the ciliate protozoa families Isotrichidae, Dasytrichidae, Entodiniinae, Diplodiniinae, Ophryoscolecinae [13], are involved in the degradation of

plant cell walls. Other important but still neglected members of the rumen microbiome are bacterial phages, which mainly include the Siphoviridae and Podoviridae families, members of the viral order Caudovirales that infect *Bacteroides, Ruminococcus*, and *Streptococcus* bacterial genera [14].

Beef cattle rumen microbiome development

Microbial colonization of the rumen starts right after birth [15]. First, a rapid invasion of aerobic bacteria occurs, followed by an increasing dominance of anaerobic taxa within days [16]. Minutes after birth, some genera such as *Methanobrevibacter* spp. and *Geobacter* spp., and some species such as *F. succinogenes*, *R. flavefaciens*, and *P. ruminicola*, are found in the rumen and gastrointestinal tract of calves [17].

A study found that the Bacteroidetes and Proteobacteria phyla decreased, while the Firmicutes phylum increased to dominate the rumen microbiome at day 14, within the first 2 weeks of life of beef calves before weaning [18]. The genera *Bifidobacterium*, *Veillonella* and *Megasphaera* increased, while the genera *Actinomycetes*, *Oscillospira*, and *Fusobacterium* decreased in the same time frame. After day 21, the microbial community did not experience drastic changes, however, the microbial alpha diversity continued to increase. The Cyanobacteria, Elusimicrobia, WPS-2 and planctomycetes phyla increased after day 21, and the genus *Fibrobacter* increased while the genera *Succinivibrio* and *Catenibacterium* decreased by day 96. Overall, the bacterial genera *Prevotella* and the archaeal genera *Methanobrevibacter* spp. were the most abundant taxa present in the rumen during early life [18]. Other study found that while calves were kept with their dams, the rumen microbiome beta diversity of both were very similar [19].

After weaning, when calves directly transitioned to a high-grain diet, the Bacteroidetes phylum increased, while Verrucomicrobia and Planctomycetes phyla and the archaeal population (mostly *Methanobrevibacter* genera) decreased. These changes were reflected in alpha diversity decrease and an increasingly dissimilar beta diversity compared to their dams [19]. When calves were supplemented with concentrate mostly composed of corn, soybean meal and molasses, the genera *Firmicutes* and *Prevotella* increased in abundance, while the genus *Fibrobacteres* decreased [20]. When creep feeding was provided to calves before weaning, the genus *Prevotella* increased

compared to calves without creep feeding, but within 4 weeks the rumen microbiome stabilized with similar composition between the groups [21]

Before the feedlot phase of beef production, the use of perennial pasture for backgrounding favored the dominance of orders Clostridiales and Bacteroidales to tackle the high fiber content of the diet [22]. When a high roughage haylage-based diet was provided for backgrounding, calves showed a substantial increase in the *Succinivibrionaceae* family, which antagonizes methanogenic bacteria in the rumen by competing for hydrogen to favor propionate production. Nonetheless, the metabolic products of *Succinivibrionaceae* family may inhibit the growth of some butyrogenic taxa, resulting in a potential imbalance in the rumen. Calves under perennial pasture backgrounding exhibited compensatory growth, improving their performance in the feedlot [22].

Rumen microbes associated with production traits in beef cattle

Many lines of evidence have associated the rumen microbiome of dairy cattle with important economic and environmental traits such as feed efficiency [23], methane emission [24] and milk composition [25,26]. However, as the purpose of dairy and beef cattle are different, their production lifecycles, genetics, nutritional needs, and relevant production traits are different as consequence. We will focus this section only on studies that associate rumen microbiome with beef cattle production traits.

Feed efficiency is one of the most valuable traits of beef cattle. It is not only oriented to reduce feed costs, which are estimated at ~70% of total input costs in beef operations [8] but also to reduce the environmental impact of the beef industry. Since the ruminal production of methane can decrease between 2 to 12% of feed efficiency in cattle [8], it is expected that efficient beef cattle will produce less methane while also reducing land use for pasture [27]. Although feed efficiency can be improved by increasing concentrate and decreasing forage in diet, it can also generate distress in the rumen leading to an adverse health impact on the animal [10]. The rumen microbiome may present some alternatives to improve feed efficiency in beef cattle.

The genus *Eubacterium*, a butyrate- and propionate- producing taxon, was positively associated with highly efficient beef cattle, while low feed efficiency was associated with some potential pathogenic members of the Proteobacteria phylum and the Spirochaetales order, which are known to produce less energy-rich substrates such as acetate [10]. A particular species of the *Eubacterium* genus, *E. ruminantium*, was found in higher abundance in efficient beef cattle, along with *Fibrobacter succinogenes* and *Megasphaera elsdenii*, [28]. Additionally, the families *Prevotellaceae, Paraprevotellaceae, Fibrobacteraceae, RF16*, *RFP12*, *S24-7, Lentisphaerae*, and *Tenericutes* were associated with selective absorption of VFAs in the rumen, which increase energy availability to the host [29]. On the other hand, a high ratio of Proteobacteria to (Firmicutes + Bacteroidetes) was found to be an indicator of rumen dysbiosis. An excessive abundance of Proteobacteria has been associated with high propionate concentration and low rumen pH [30].

Regarding methane emissions, there is a clear positive correlation between *Methanobrevibacter* and *Methanosphaera* abundance and ruminal methanogenesis in beef cattle [31]. Specifically *M. ruminantium* has been associated with high methane emitter cattle [32]. In addition, the *Butyrivibrio* and *Pseudobutyrivibrio* genera were highly correlated with high methane emissions [33], as well as the genus *Desulfovibrio* [31]. Conversely, the genera *Methylomonas, Megasphaera* and *Acidaminococcus* [33] were associated with low methane emissions, as well as the *Succinivibrionaceae* family [31], which may be associated with the competition between this taxon and methanogens for hydrogen in the rumen [34].

Meat quality is an important trait for beef cattle production. The Verrucomicrobia phylum, the Oscillospira class, families RFP12 and Porphyromonadaceae, and the *Paludibacter* genus were positively correlated with high marbling score (intermuscular fat deposition), whereas the *Olsenella* genus was negatively correlated [35]. Another study found that the family S24-7 is positively correlated with marbling score and *longissimus* muscle lipid content [36]. Interestingly, this family was also positively correlated with increased intramuscular fat in pigs [37]. Also, some members of the Proteobacteria phylum in the rumen microbiome have been positively associated with an increased beneficial fatty acid content in beef, such as the *Pseudomonas*, *Vibrio*, *Aeromonas* and *Serratia* genera.

These taxa have been shown to have lipolytic activity and have been involved in microbial protein synthesis, carbohydrate metabolism and transport, and lipopolysaccharide biosynthesis [38].

It is important to mention that some of these beef cattle rumen microbiome taxa have some degree of heritability. Four key bacterial taxa were found to be heritable and highly correlated with other bacteria in the beef cattle rumen: unclassified Clostridiales, unclassified Succinivibrionaceae, unclassified Coriobacteriaceae, and unclassified Christensenellaceae. The phylum Bacteroidetes showed low heritability, while the Firmicutes phylum showed moderate heritability [39]. Although not a production trait, but relevant for public health, there is evidence that the species *Ruminococcus* spp, *Prevotella ruminicola*, *Muribaculaceae* spp. and *Collinsella aerofaciens* are bacterial hosts of highly expressed antimicrobial resistance genes in the rumen of beef cattle [40].

SEQUENCING METHODS TO STUDY THE RUMEN MICROBIOME OF BEEF CALVES

As shown above, several studies have explored the rumen microbiome of beef cattle, however most of these studies have not extensively addressed the benefits and limitations of the sequencing method used. Microbiome research, in general, is based on bulk DNA sequencing through long and complex protocols, from sample collection to bioinformatic analysis [41]. Each point within this workflow has its own challenges and limitations, which oftentimes cannot be tackled due to budgetary and technical constraints.

16S rRNA gene sequencing was the first method used to study microbial communities in the rumen [42], as it offers a cost-efficient taxonomic description of bacterial and archaeal microorganisms. The method is based on the amplification of a hypervariable region of the 16S rRNA gene, a well-conserved gene in bacteria and archaea [43]. An extensive explanation of the mechanism, advantages and limitations of the method is offered in chapter 03. However, it is important to mention that 71% of the studies focused on the rumen microbiome of beef cattle listed in this literature review used 16S rRNA gene sequencing **(Table 1.2).**

Shotgun metagenomic sequencing is an untargeted method that sequences all the DNA content in a sample [44]. Despite the abundant information that this method provides about a microbial community, it is still cost- and resource- prohibitive for many rumen microbiome researchers. A detailed explanation of the method can be found in chapter 03. As DNA sequencing price steadily decreases, shotgun metagenomic sequencing gains more popularity in the rumen microbiome research field. Specifically in beef cattle rumen microbiome research, 29% of studies listed in this literature review used shotgun metagenomic sequencing **(Table 1.2).**

The DNA sequencing reads obtained by shotgun metagenomic sequencing are classified using reference database [44]. However, it has been evidenced that a high proportion of sequencing reads obtained from rumen microbiome samples are not classifiable [13,45–48]. This relevant challenge needs further investigation because many reported associations in literature may be based on misclassification and may not reflect an accurate diversity of microorganisms. Since both the shotgun metagenomic and 16S rRNA methods rely heavily on reference databases, the lack of well-described genomes of the rumen microbial community impedes a complete classification of the DNA obtained from the rumen, particularly for metagenomic studies [46–48]. Researchers have attempted to expand reference databases with genomes obtained from rumen samples, and this effort did improve classification rates in some studies [45,47,48]. However, even with these efforts, between 30 to 50% of rumen metagenomic data remain unclassified. Some ways to address this issue are discussed in chapter 02 and chapter 03 of this manuscript.

Based on our literature review, we identified only a single study that used both16S rRNA gene and shotgun metagenomic sequencing of rumen microbiome samples; and this study involved dairy cattle. Although both methods were used, a formal comparison of the methods was not explicitly explored in the study. Overall, shotgun metagenomic sequencing identified more genera than 16S rRNA sequencing, and discrepancies in the microbial abundance of the phyla Firmicutes and Bacteroidetes, and the genus *Selenomonas* were reported [49].

Within the cattle rumen microbiome research field (both dairy and beef), 16S rRNA gene sequencing is still commonly used. In a PubMed literature search, 91.5% (279/305) of studies that contain the research query “(cattle) AND (rumen) AND (microbiome)” used 16S as the sequencing method, while 8.5% (26/305) with the same research query used shotgun metagenomics (accessed in May 2023). A growing number of studies infer important associations of a given relevant outcome with rumen microbiome features without further phenotypic or mechanistic validation. Also concerning is the prospect that the inconsistent and still changing methodological approach to study the rumen microbiome may present incompatibility issues for robust comparisons between studies, both past and future. Both 16S rRNA and shotgun metagenomic sequencing produce valuable information about the microbial community of the rumen, but as DNA sequencing price decreases, shotgun metagenomic sequencing will likely become a commonly used method. In this context, a comparison of performance between 16S rRNA gene sequencing and shotgun metagenomic sequencing is necessary to understand the benefits and limitations of each method, and even more importantly how each method may generate systematically biased results. Understanding methodological sources of bias is critical for interpreting the results from beef cattle rumen microbiome studies.

THE IMPACT OF EARLY-LIFE MANAGEMENT PRACTICES ON THE RUMEN MICROBIOME OF BEEF CATTLE

The early life of cattle is a key stage for rumen microbiome development, as the assembly of the initial microbial community in the rumen has a long-lasting impact [15]. Correct priming of the rumen microbiome early in the life of cattle may lead to a predictable mature microbiome state [16]. Host genetics and diet are deterministic factors that govern the development of a microbial community in the rumen. Increasing evidence shows that the rumen microbiome is subjected to host genetic control to some extent [50], as it has a core microbiome that is heritable and quite stable later in adult life [9,16]. The dietary regimen in early life might be another potential way to select a given production trait during animal development, particularly because the rumen is quite dynamic and plastic at this stage of development [15]. Later, as the cattle age to

maturity, dietary interventions can have a significant but not permanent effect on the rumen microbiome [7].

Some studies showed that a higher concentration of dietary fiber had some beneficial outcomes compared to a higher concentrate content, but also may have some important pitfalls. A forage-based diet in beef cattle favors the abundance of Planctomycetes, LD1-PB3 and SR1 phyla and *Lachnospira* and *Sutterella* genera in the rumen, while a grain-based diet favors the abundance of starch-fermenting bacteria of the family of *Succinivibionaceae* and the genus *Succinimonas* [51]. Likewise, a dry lot backgrounding (grain-based) diet for beef calves favored the abundance of the *Succinivibrionaceae* family. This taxon antagonizes methanogenic bacteria but may also inhibit the growth of some butyrogenic bacteria, which can potentially have detrimental effects on rumen health [22]. Importantly, forage-based diets increase the rumen microbiome alpha diversity but also increase methanogenic bacteria, as they have been shown to be positively correlated [8]. Thus, dietary requirements may typically meet a delicate grain-forage balance, and altering this balance can lead to unintended or negative consequences.

The impact of weaning has been broadly described in dairy cattle. Calves weaned at 48 days of life had a reduction in alpha diversity indices and a shift in beta diversity post-weaning, mainly led by an increase of the Proteobacteria and Firmicutes phyla and the *Prevotella* genus, and simultaneous decrease of the *Bacteroides* genus. However, weaning strategy (i.e., gradual or abrupt) had a limited impact on the development of the rumen microbiome of weaned dairy calves [52]. Importantly, the age of weaning has been shown to have an impact on the maturation process of the rumen. Early-weaned calves, either at 6 [53] or 7 weeks of age [54], have shown reduced growth rate, reduced alpha diversity and a rapid shift of rumen microbiome beta diversity compared to late-weaned calves. Late-weaned calves, either at 8 [53] or 48 weeks of age [54], showed a gradual beta diversity change in the rumen microbiome, resulting in less ruminal distress [53]. A decrease in the genus *Olsenella*, a carbohydrate fermenter, and an increase in unclassified *Clostridia* and *Fibrobacteres* genera were observed post-weaning as a higher forage content was introduced to the calves' diet [54].

The effect of weaning has not been formally studied in beef calves. However, using 16S rRNA gene sequencing, a series of studies have described the temporal dynamics of the rumen microbiome of beef calves from birth to weaning [55], from day 7 to day 96 of life [18] and from weaning day to feedlot arrival [19]. Importantly, one study assessed the effect of different creep-feeding diets before weaning compared to a uniform post-weaning diet in beef calves [21]. The temporal dynamics of beef cattle rumen microbiome after weaning, and once they received a concentrate-based diet (high grain content) in the feedlot, showed an increase of the Bacteroidetes phylum, and a decrease of the Verrucomicrobiota and Planctomycetes phyla and the *Methanobrevibacter* genus. These changes were reflected in an alpha diversity decrease post-weaning, which was maintained at a decreased level until arrival at the abattoir. Notably, the beta diversity of the calves' rumen microbiome at weaning was similar to their dams, but dissimilarity increased once they moved to a feedlot [19].

Although a common practice in beef calf management, the effect of castration on the rumen microbiome has not been formally assessed. However, in dairy cattle, castration at 36 weeks of age resulted in an alpha diversity decrease and an increase in the abundance of the *Prevotellaceae* family in the rumen 3 days after the procedure, which was more pronounced in low-weight calves. Interestingly, members of the *Prevotellaceae* family have been previously associated with stressful procedures, such as heat stress [56]. Furthermore, the authors pointed to a masking effect of age, which may confound the association between rumen microbiome changes and the stress produced by invasive procedures performed at an early age [57].

Although valuable for cattle research in general, most of what we know pertains to dairy cattle research. Since the production lifecycle and nutritional requirements of beef cattle are different, the trajectory of rumen microbiome maturation may differ from dairy cattle, particularly because beef cattle experience multi-factor changes at each production stage (e.g., cow-calf, backgrounding, feedlot). Lack of longitudinal studies in beef cattle hinders a complete picture of the rumen microbiome development, especially during early life events. What is more, the impact of common early-life management practices, such as castration timing and weaning strategy, on the rumen microbiome of beef cattle

has not been explored through a comprehensive metagenomic sequencing approach. Most of the works cited in this literature review have used 16S rRNA gene sequencing to study the beef cattle rumen microbiome, which precludes the investigation of important taxa other than bacteria and archaea, and the study of the putative function of the rumen microbial community. As microbiome research moves to shotgun metagenomic sequencing, comparing both sequencing methods will help to integrate and fully leverage extant 16S data to understand the beef cattle rumen microbiome.

TABLES AND FIGURES

Table 1.1. Relevant microorganisms in cattle rumen.

Taxonomic rank	Name	Function	Reference
Order			
	Bacteroidales	Part of the core microbiome	[9]
	Clostridiales	Part of the core microbiome	[9]
Family			
	Lachnospiraceae	Part of the core microbiome	[9]
	Paraprevotellaceae	Part of the core microbiome	[9]
	Prevotellaceae	Part of the core microbiome	[9]
	RF16	Part of the core microbiome	[9]
	RFP12	Part of the core microbiome	[9]
	S24−7	Part of the core microbiome	[9]
	Succinivibrionaceae	Part of the core and heritable microbiome, predominant family, cellulolysis, competes which methanogens for hydrogen witch this family use to produce succinate, potential target for methane emission mitigation	[6,9]
	Victivallaceae	Part of the core microbiome	[9]
Genus			
	Acidaminococcus	Associated with low feed efficiency	[10]
	Anaeromyces	Predominant fungi	[7]
	Anaeroplasma	Part of the core microbiome	[9]
	BF311	Part of the core microbiome	[9]
	Butyrivibrio	Part of the core microbiome, predominant genus, degrades hemicellulose and metabolizes polysaccharides and peptides	[6,9]
	Buwchfawromyces	Predominant fungi	[7]
	Caecomyces	Predominant fungi	[7]
	Coprococcus	Part of the core microbiome, involved in methanogenesis and acetogenesis, produce lactate and acetate, use hydrogen and lactate to produce propionate via acrylate metabolic pathway	[6]
	Cyllamyces	Predominant fungi	[7]
	Desulfococcus	Associated with low feed efficiency, reduces sulfate	[10]
	Entodinium	Predominant protoza	[7]
	Epidinium	Predominant protoza	[7]
	Eubacterium	Associated with high feed efficiency	[10]
	Eudiplodinium	Predominant protoza	[7]
	Fibrobacter	Predominant genus, metabolizes polysaccharides and peptides	[7]
	Lactobacillus	Associated with high feed efficiency	[10]

Table 1.1. (continued).

Taxonomic rank	Name	Function	Reference
Genus			
	Megasphaera	Competes with methanogens for hydrogen which this genus use to produce propionate via acrylate metabolic pathway, potential target for methane emission mitigation	[6]
	Methanobrevibacter	Predominant archaea that account for up to 70% of rumen archaeal community, positively correlated with methane emissions	[6]
	Methanosphaera	Part of the core microbiome, positively correlated with methane emissions	[6]
	Neocallimastix	Part of the core microbiome, predominant fungi	[7,9]
	Ontomyces	Predominant fungi	[7]
	Orpinomyces	Predominant fungi	[7]
	Piromyces	Predominant fungi	[7]
	Polyplastron	Predominant protoza	[7]
	Prevotella	Part of the core microbiome, predominant genus, degrades hemicellulose and metabolizes polysaccharides and peptides	[6,9]
	Providencia	Associated with low feed efficiency	[10]
	Pseudobutyrivibrio	Predominant genus, degrades non-crystalline hemicellulose	[6]
	Ruminococcus	Part of the core microbiome, cellulolysis	[9]
	Sharpea	Competes with methanogens for hydrogen which this genus use to produce lactate, potential target for methane emission mitigation	[6]
	Sphaerochaeta	Associated with low feed efficiency	[10]
	Succiniclasticum	Associated with low feed efficiency	[10]
	Treponema	Associated with low feed efficiency	[10]
	Trichomonas	Protozoa associated with low feed efficiency	[10]
	Variovorax	Associated with low feed efficiency, reduces sulfate	[10]
Species			
	Acetitomaculum ruminis	Part of the core microbiome, ferments soluble sugar to produce acetate, propionate and butyrate, involved in methanogenesis and acetogenesis	[6]
	Anaeroplasma abactoclasticum	Part of the core microbiome, ferments soluble sugar to produce acetate and butyrate	[6]
	Bifidobacterium ruminale	Part of the core microbiome, ferments soluble sugar to produce acetate and lactate	[6]
	Butyrivibrio fibrisolvens	Part of the core microbiome, degrades crystalline cellulose and hemicellulose (polymers), ferments soluble sugar to produce acetate and butyrate	[6]
	Butyrivibrio hungatei	Part of the core microbiome, degrades crystalline cellulose and hemicellulose (polymers), ferments soluble sugar to produce acetate and butyrate	[6]
	Butyrivibrio proteoclasticus	Part of the core microbiome, degrades crystalline cellulose and hemicellulose (polymers), ferments soluble sugar to produce acetate and butyrate	[6]
	Christensenella minuta	Part of the core microbiome, ferments soluble sugar to produce succinate, lactate, acetate and butyrate	[6]

Table 1.1. (continued).

Taxonomic rank	Name	Function	Reference
	Fibrobacter succinogenes	Part of the core microbiome, degrades crystalline cellulose and hemicellulose (polymers), ferments soluble sugar to produce succinate, acetate and formate	[6,9]
	Lachnoclostridium clostridioforme	Part of the core microbiome, ferments soluble sugar to produce acetate and lactate	[6]
	Lachnospira multiparus	Part of the core microbiome, ferments soluble sugar to produce acetate, lactate and formate	[6]
	Methanobrevibacter gottschalkii	Predominant archaea, involved in methanogenesis	[7]
	Methanobrevibacter ruminantium	Predominant archaea, involved in methanogenesis and acetogenesis	[6,7]
	Methanosphaera stadtmaniae	Predominant archaea, involved in methanogenesis and acetogenesis	[6]
	Olsenella umbonata	Part of the core microbiome, ferments soluble sugar to produce acetate and lactate	[6]
	Oribacterium sp. strain C9	Part of the core microbiome, ferments soluble sugar to produce acetate and lactate	[6]
	Prevotella ruminicola	Part of the core microbiome, degrades crystalline cellulose and hemicellulose (polymers), ferments soluble sugar to produce acetate, lactate, succinate, propionate and formate	[6]
	Pseudobutyrivibrio ruminis	Part of the core microbiome, ferments soluble sugar to produce acetate	[6]
	Pseudobutyrivibrio xylanivorans	Part of the core microbiome, degrades crystalline cellulose and hemicellulose (polymers), ferments soluble sugar to produce acetate and butyrate	[6]
	Pseudoscardovia suis	Part of the core microbiome, ferments soluble sugar to produce acetate and lactate	[6]
	Roseburia faecis	Part of the core microbiome	[9]
	Ruminobacter amylophilus	Part of the core microbiome, ferments soluble sugar to produce acetate, lactate, succinate and formate	[6]
	Ruminococcus albus	Part of the core microbiome, predominant specie, degrades cellulose	[6,9]
	Ruminococcus flavefaciens	Part of the core microbiome, degrades crystalline cellulose and hemicellulose (polymers), ferments soluble sugar to produce succinate, acetate and formate	[6,9]
	Selenomonas ruminantium	Part of the core microbiome, predominant species, degrades crystalline cellulose and hemicellulose (polymers), ferments soluble sugar to produce succinate, acetate, lactate, propionate and formate	[6]
	Succiniclasticum ruminis	Part of the core microbiome, ferments soluble sugar to produce acetate, succinate and propionate, use secondary fermentation products	[6]
	Succinimonas amylolytica	Part of the core microbiome, ferments soluble sugar to produce acetate, lactate, succinate and formate	[6]
	Succinivibrio dextrinosolvens	Part of the core microbiome, ferments soluble sugar to produce acetate, lactate, succinate and formate	[6]
	Treponema bryantii	Part of the core microbiome, predominant specie, ferments soluble sugar to produce acetate and succinate	[6]

Table 1.2. Beef cattle rumen microbiome studies listed in this literature review.

Method	Study design	Sample size	Country	Year	Beef cattle age	Beef cattle breed	Reference
16S-V1-V3 & 16S-V6-V8	Cross-sectional	709	China	2019	Adult	Purebred Angus, Charolais , and the Kinsella composite hybrid	[39]
16S-V3-V4	Cross-sectional	6	Brazil	2021	Adult	Nellore pure breed	[58]
16S-V3-V4	Cross-sectional	14	Korea	2020	Adult	Korean beef cattle (Hanwoo)	[35]
16S-V3-V4	Cross-sectional	18	USA	2022	Adult	Wye Angus beef	[51]
16S-V3-V4	Cross-sectional	6	USA	2020	Calves (155 d)	Angus	[20]
16S-V3-V4	Longitudinal	12 (x4)	USA	2022	Post-weaned Calves (7.5 m) to abattoir	Angus	[19]
16S-V3-V4	Cross-sectional	24	USA	2020	Adult	Angus	[36]
16S-V3-V4	Cross-sectional	25	USA	2019	Calves pre and post weaning	Angus	[21]
16S-V4	Cross-sectional	586	USA	2020	Adult	Cross bred	[29]
16S-V4	Cross-sectional longitudinal	38	USA	2022	Post-weaned calves	Angus and Angus x Simmental	[22]
16S-V4	Cross-sectional	247	USA	2018	Adult	Cross bred	[59]
16S-V4 & 16S-V6-V8	Cross-sectional	35	Ireland	2020	Calves (7d-96d)	Aberdeen Angus-sired	[18]
Shotgun	Cross-sectional	72	UK	2020	Adult	Purebred Luing and crossbred Charolais	[10]
Shotgun	Cross-sectional	50	UK	2018	Adult	Purebred Luing (LU), crossbred Charolais (CH, Aberdeen Angus (AA) and Limousin (LIM)	[33]
Shotgun	Cross-sectional	48	USA	2022	Adult	Angus, Charolais, Kinsella composite hybrid	[40]
Shotgun	Cross-sectional	363	UK	2022	Adult	Cross bred of Aberdeen Angus and Limousin breeds, Charolais-crosses and purebred Luing	[38]
Shotgun	Cross-sectional	72	UK	2015	Adult	Aberdeen Angus and Limousin cross bred	[31]

OBJECTIVES

Overall aim

Study the effect of two early-life common management practices on the rumen metagenome of beef cattle through a comprehensive metagenomic sequencing approach.

Specific aims

SA 1: Investigate the effect of four castration timing windows and two weaning strategies on the rumen metagenome of beef calves.

SA 2: Compare the performance of shotgun metagenomic sequencing and 16S rRNA gene hypervariable region V4 sequencing for taxonomic characterization of beef calf rumen microbiome.

CHAPTER 02: EFFECT OF CASTRATION TIMING AND WEANING STRATEGY ON THE RUMEN METAGENOME OF BEEF CALVES

Gerardo R. Diaz[1], Tara N. Gaire[1], Peter Ferm[1], Lacey Casey[1], Luciano S. Caixeta[1], Timothy J. Goldsmith[1], Joe Armstrong[1], Noelle R. Noyes[1]*

[1] Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, 55108, USA

* Corresponding author: Noelle R. Noyes (nnoyes@umn.edu)

Authors' email addresses:

Gerardo R. Diaz diazo005@umn.edu

Tara N. Gaire tgaire@umn.edu

Peter Ferm fermx014@umn.edu

Luciano S. Caixeta lcaixeta@umn.edu

Timothy J. Goldsmith gold0188@umn.edu

Joe Armstrong armst225@umn.edu

This chapter is in preparation to be submitted to a peer-review journal

ABSTRACT

Background

Beef cattle experience several management challenges across their lifecycle. Castration and weaning, two major interventions in early life of beef cattle, have demonstrated a substantial impact on animal performance. Despite the key role of the rumen microbiome on productive traits of beef cattle, the effect of castration timing and weaning strategy on this microbial community has not been formally described. We assessed the effect of four castration time windows (at birth, at turnout, pre-weaning and at weaning) and two weaning strategies (fence-line and truck transportation) on the rumen metagenome in a randomized controlled study with 32 male calves across 3 collection days. Ruminal fluid samples were submitted to shotgun metagenomic sequencing and changes in the taxonomic and functional profile of the rumen metagenome were described.

Results

Using a comprehensive yet stringent taxonomic classification approach, we identified 37,894 unique taxa classified under 7 archaeal, 40 bacterial, 14 viral, 6 protozoa, 3 fungal phyla across all samples. Castration timing had a limited long-term impact in the rumen metagenome and was not associated with changes in alpha and beta diversity. The interaction of collection day and weaning strategy was associated with changes in the rumen microbiome, having a significant decrease in alpha diversity and differences in beta diversity within 48 hours post-weaning, especially in calves abruptly weaned by truck transportation. Calves sequentially weaned using a fence-line weaning strategy had lower relative abundance of Bacteroides, Lachnospira, Fibrobacter and Ruminococcus compared to calves weaned by truck transportation. While some genes (DNA level) from the hydrogenotrophic methanogenesis pathway (fwdB and fwdF) had higher relative abundance in fence-line-weaned calves post-weaning, the antimicrobial resistance gene tetW consistently represented more than 50% of the resistome across time, weaning and castration groups, without any change in the relative abundance.

Conclusions

Within the context of this study, castration timing had limited long-term effects on the rumen metagenome while weaning strategy had short-term effects on the rumen metagenome. These effects influenced the microbiome composition and methanogenic genes but not the rumen resistome.

Keywords: Cattle, early life, Microbiome, Methanogens, Resistome, Fence line, truck, Gastrointestinal, Shotgun

BACKGROUND

Beef cattle, like other ruminants, have the ability to turn non-edible feedstuff into highly nutritious resources that humans could not otherwise use [6]. The microbial community inhabiting the rumen, termed the rumen microbiome, achieves this through fermentation processes that can generate up to 70% of animal energy needs [8]. These processes also result in the generation of greenhouse gases (e.g., methane), which have been estimated to contribute up to 40% of total livestock emissions worldwide [60]. Recent culture-independent molecular techniques have allowed an intensive exploration of the rumen microbiome and its association with relevant production traits in the bovine host. Multiple members of this microbial community, made up of bacteria, fungi, archaea, protozoa and viruses [7], have been associated with feed efficiency [10,23,59], methane emissions [24,33,61] and meat quality [38]. The bacterial and archaeal communities of the rumen microbiome have been shown to have a life-long influence on not only their host, but also that host's offspring [9,39]. For example, the initial colonization of microbes in the neonatal calf rumen is a significant predictor of rumen microbiome composition later in life, i.e., once the rumen has developed [15], and a small bacterial subset of the rumen microbiome has been identified as a heritable microbial core in adult cattle [9]. In light of current evidence, manipulation of the rumen microbiome may be an effective strategy to improve many aspects of beef production, much like leveraging cattle genetics has led to myriad impacts across the beef production system. However, we still lack longitudinal studies to understand how external factors impact the establishment, composition and function of the rumen microbiome throughout the beef cattle life cycle [6,18].

A typical beef cattle lifecycle can involve moving animals through several stages, including cow-calf, backgrounding, stocker-feeder, and feedlot [4]. As calves move through these stages, they experience different management practices that can include various physical interventions (e.g., castration, weaning, dehorning); changes in diet, environment and herd; as well as transport [62]. Some of these changes happen suddenly and represent stressors that can affect a calf's metabolism, immune system, health and performance [63]. Specific evidence shows that choice of strategy for

castration [64], weaning [65], dehorning [66] and high-energy diet supplementation [67] can directly impact welfare, health and performance of beef cattle.

Similarly, a growing body of evidence suggests that the rumen microbiome of beef cattle is also impacted by these management practices, including diet changes [20,22,68–70] and stress factors [58]. Despite the importance of early-life events for both long-term beef cattle performance [3] and rumen microbiome assembly and establishment [15], very little is known about the effect of management practices on the beef calf rumen microbiome, especially in comparison to the body of literature that pertains to dairy calves [52–54,57]. Furthermore, the existing literature related to the rumen microbiome of beef calves is based on 16S rRNA gene sequencing [18,20,71] instead of shotgun metagenomic sequencing. This technical difference has important repercussions on the insight that can be gained from these studies. Specifically, the use of 16S rRNA data does not allow for detection of microbial organisms other than bacteria and archaea with a limited taxonomic resolution [26]; or for the characterization of the putative function of ruminal microbes, which play an important role in antimicrobial resistance [30], environmental impact [61] and productivity [10] of adult cattle.

While weaning and castration of beef calves have been shown to significantly impact animal performance (i.e., food intake, average daily gain, and health status), the effect of these practices on the rumen microbiome remains unexplored. We hypothesized that castration timing and weaning strategy impact the diversity, taxonomy and potential functional profile of the rumen microbiome of beef calves. To test this hypothesis, we designed a longitudinal randomized controlled study in which 32 male calves from a single cow-calf herd were randomly assigned at birth to 4 different castration windows and 2 different weaning strategies using a factorial design. Other than castration timing and weaning strategy, calves were managed identically and kept together within the same herd. Ruminal fluid was collected twice before weaning and once 48 hours after weaning **(Sup Figure 2.1).** Samples were sequenced by shotgun metagenomics and analyzed using several bioinformatic pipelines. In addition to overall taxonomic microbiome composition, we analyzed the effect of our interventions on gene composition associated with 2 relevant processes driven by microbial communities: antimicrobial resistance (AMR) and methane emissions. Regression models were used

for hypothesis testing, considering collection day, castration timing and weaning strategies as independent variables (predictors), and diversity metrics and differential abundance of a given feature as dependent variables (outcomes).

RESULTS

Low classification rate but high resolution in the taxonomic profiling of the rumen microbiome with kraken2

We collected 95 samples of rumen fluid: 32 at pre-weaning processing (September 21st), 32 at weaning (October 18th) and 31 at post-weaning (October 20th). One calf assigned to turnout castration and fence-line weaning was dropped from the study due to health issues that occurred at weaning that eventually ended in euthanasia. Despite formal randomization, there was a significant difference in age (days) and average daily gain (ADG) of calves weaned via fence-line and truck (Kruskal-Wallis Test, $P < 0.05$). Calves weaned by fence-line were on average 6 days younger and had on average 0.11 kg higher ADG compared to truck-weaned calves **(Table 2.1).**

Samples were subjected to total DNA extraction and shotgun metagenomic sequencing, which generated 5.1 x 10^9 paired-end (PE) sequencing reads across all 95 rumen fluid samples (mean 55.8 x 10^6 PE reads per sample, range 30.5 – 75.8 x 10^6) with an overall mean quality score of 35.1. The initial standard protocol for taxonomic profiling of the rumen microbiome (see methods) classified a low proportion of initial sequencing reads **(Sup. Table 2.1)**. Based on previous evidence [46], we customized our protocol for the rumen microbiome, including a host decontamination of *Bos taurus* and dietary plants sequences and the bacterial taxonomic classification using a customized database that included rumen-specific bacteria and archaea (see methods). After trimming low quality sequencing reads and removing host and dietary plants reads, an average of 38.3 x 10^6 PE reads per sample remained (70.7% of raw reads). The trimmed non-host reads were classified using Kraken2 (confidence score = 0.1) with the rumen-specific database, resulting in an average of 1.6 x 10^6 PE reads classified per sample (4.2% of trimmed non-host reads). Our customized protocol increased the rate of host-reads removal, from 22.7% using only *Bos taurus* genome to 29.3% using *Bos taurus* and dietary plants genomes; while increased the proportion of non-host reads classified by kraken2 (confidence score = 0.1), from 3.1% using a standard database to 4.2% using a rumen-specific database **(Sup. Table 2.1).** Overall, with our customized protocol, 96.2% of

classified reads were resolved to the phylum level, while 75.7% were resolved to the species level **(Sup. Table 2.2)**.

Two positive controls (mock communities) sequenced alongside the rumen samples yielded 51.6 and 47.6 x 10^6 PE reads per sample, while the 3 negative controls yielded 10298, 97 and 82 x 10^3 PE reads per sample. Positive and negative controls were classified with kraken2 using the same approach as with the rumen samples. The positive controls obtained 32.4 and 29 x 10^6 PE reads classified per sample, and the negative controls had 244, 6 and 4 x 10^3 PE reads classified per sample. At the genus level, the positive controls contained *Listeria* spp. as the predominant taxa, which was expected according to the mock community composition (ZymoBIOMICS Microbial Community Standard II Log Distribution – Catalog N° 6310). *Bacillus* spp., *Saccharomyces* spp. and *Enterococcus* spp. were above their expected abundances (0.89%, 0.89% and 0.0089%, respectively), while *Pseudomonas* spp., *E. coli* and *Salmonella* spp. were under their expected abundance (8.9%, 0.089% and 0.089%, respectively). The negative controls contained mostly *Butyrivibrio* spp., *Prevotella* spp., *Cutibacterium* spp. and human DNA, which are expected contaminants from rumen samples and human manipulation **(Sup. Figure 2.2).**

Rumen microbiome of beef calves was depicted as a complex and dynamic community by kraken2 using a comprehensive reference database

A total of 37,894 OTUs (operational taxonomic units) were identified across the 95 rumen fluid samples by kraken2 using a confidence score = 0.1 and a comprehensive reference database that included archaea, bacteria, virus, plasmids, human, UniVec_core, protozoa and fungi reference genomes, and 410 rumen-specific bacteria genomes from the Hungate project [72]. We identified OTUs at every taxonomic level for Bacteria, Archaea, Fungi, Protozoa and Virus, but we focused our downstream analysis on the bacterial and archaeal domains due to their major role in ruminant metabolism. Within the bacterial domain (98.6% of the total classified reads), 40 unique phyla, 93 classes, 219 orders, 519 families, 1909 genera and 8674 species were identified across all rumen fluid samples. Within the archaeal domain (1.4% of the total classified reads), 7 unique phyla, 18 classes, 31 orders, 48 families, 140 genera and 351 species were identified **(Sup. Table 2.3)**. Overall, the most abundant bacterial phyla across all ruminal

fluid samples were *Bacillota* (48.1% ± 13%, mean ± SD), *Bacteroidota* (42.7% ± 11%), *Fibrobacteres* (5.16% ± 3.48)*, Pseudomonadota* (2.05% ± 0.62)*, Actinomycetota* (1.57% ± 0.36) *and Spirochaetes* (1.28% ± 0.24) **(Figure 2.1.A)**. The 10 most abundant bacterial genera across all ruminal fluid samples were *Prevotella* sp. (40.5% ± 10.9, mean ± SD), *Butyrivibrio* sp. (16.4% ± 7.72), *Selenomonas* sp. (5.50% ± 2.64), *Fibrobacter* sp. (5.46% ± 3.67), *Oribacterium* sp. (3.9% ± 1.31), *Succiniclasticum* sp. (3.73% ± 2), *Pseudobutyrivibrio* sp. (3.65% ± 1.73), *Ruminococcus* sp. (2.94% ± 1.28), *Eubacterium* sp. (1.8% ± 0.96), and *Sarcina* sp. (1.82% ± 0.78). Within the archaeal domain across all ruminal fluid samples, the 5 most abundant genera were *Methanobrevibacter* sp. (83.8% ± 7.03), *Methanosphaera* sp. (14.7% ± 4.04), *Methanomicrobium* sp. (2.72% ± 2.49), *Methanosarcina* sp. (1.55 ± 0.1) and *Candidatus Methanoplasma* sp. (1.45% ± 0.39) **(Sup. Figure 2.3).**

Alpha diversity results were summarized using 2 main indices at the genus level: richness and Shannon's index. We used linear mixed effects modeling with indices as response (outcome) variable; and collection day, castration timing and weaning strategy as explanatory (predictor) variables. We added age and average daily gain (ADG) as potential confounders, and animal ID as a random effect to account for non-independence due to repeated measures on each calf. Collection day was significantly associated *(ANOVA-III P*<0.001) with Shannon's index, but not with richness **(Figure 2.2.A)**. Specifically, rumen samples collected at post-weaning contained significantly lower Shannon's diversity (adjusted mean ± SE, 2.18 ± 0.03) than before weaning (2.50 ± 0.03) **(Figure 2.2.B)**.

The rumen microbiome composition shifted significantly in the 48 hours post-weaning. Using multivariate zero-inflated Gaussian mixture models with animal ID as a random effect and collection day, weaning strategy, age and average daily gain (ADG) as fixed effects, we measured the differential abundance (log_2 fold change) of phyla between collection days, weaning and castration groups. We observed numerous differentially abundant phyla when comparing samples collected at weaning and post-weaning. Eight phyla were in higher abundance at post-weaning, while seven phyla were in lower abundance. Considering only phyla with high average abundance across samples, the phyla *Spirochaetes, Euryarchaeota* and *Bacillota* were in lower abundance post-weaning

compared to at-weaning, while phyla *Fibrobacteres* and *Bacteridota* were in higher abundance **(Figure 2.2.C).**

Dissimilarity matrices of Bray-Curtis distances at the genus level were used for non-metric multidimensional scaling (NMDS) ordination and permutational multivariate analysis of variance (PERMANOVA) testing of rumen microbiome variation by collection day, castration timing, and weaning strategy; the marginal effect (R^2) was obtained for each predictor (option by="margins" in *adonis2*). A relatively large and statistically significant proportion of between-sample rumen microbiome variability was partitioned to collection day ($R^2 = 31.9$, $P < 0.001$) **(Figure 2.3.B)**.

Weaning strategy but not castration timing was associated with rumen microbiome differences in beef calves

Not only collection day but also its interaction term with weaning strategy was significantly associated with alpha diversity of the rumen microbiome (*ANOVA-III, P<0.001*). Specifically, rumen samples collected from truck-weaned calves 48 hours after weaning had significantly lower diversity compared to the fence-line weaned calves sampled at pre-weaning (β = -0.18, 95% CI = -0.31, -0.05) **(Figure 2.3.A).** Weaning strategy was associated with 5.2% (PERMANOVA $P<0.01$) of the overall variability in beta diversity across all of the rumen samples **(Figure 2.3.B)**. However, when only post-weaning samples were analyzed, weaning strategy was associated with 32.5% (PERMANOVA $P<0.01$) of the between-sample variability.

Differential abundance testing by weaning strategy stratified by collection day showed that the majority of differences between weaning groups were observed at the post-weaning time point. Specifically, 564 genera identified in the post-weaning samples were significantly differentially abundant between fence-line and truck-weaned calves, compared to 266 and 239 genera at the pre-weaning and weaning collection days, respectively **(Figure 2.3.C).** Considering only the most abundant genera across all samples, we identified *Bacteroides*, *Lachnospira*, *Petrimonas, Micromonospora, Fibrobacter, Sarcina, Streptococcus* and *Ruminococcus* genera as having significantly lower abundance (log2 fold change < -1) in fence-line-weaned calves compared to truck-weaned calves post-weaning; while the *Rhodococcus, Agrobacterium, Anaerovibrio,*

Oribacterium, Plantibacter, Variovoraxgenera and *Lachnoclostridium* genera were in higher abundance (log2 fold change > 2) **(Sup. Table 2.4).**

Unlike weaning strategy, castration timing was not significantly associated with differences in Shannon's index at any collection day (pre-weaning *P*=0.4, weaning *P*=0.8 and post-weaning *P*=0.8). When beta diversity was assessed separately for each collection day, the variation partitioned to castration timing was small and not statistically significant **(Figure 2.4.A - C)**. These results do not provide evidence that castration timing impacts the diversity of the rumen microbiome of beef calves, at least during the collection time points included in this study. Differential abundance testing revealed limited statistically significant differences in relative abundance of any phyla when comparing castration timing groups at pre-weaning, weaning (data not shown) and post-weaning **(Figure 2.4.D).**

Weaning strategy was associated with differences in the relative abundance of methanogenic microbes and genes

We assessed the effect of weaning strategy specifically on the methanogenic microbial community and associated genes using a dedicated database and bioinformatic tool [73]. The relative abundance of these microbes across time and between weaning strategies was highly heterogeneous **(Sup Figure 2.4.A.)**. As with the overall microbiome, the associations between methanogenic gene diversity and both weaning strategy and time were significant. Shannon's index for methanogenic genes was significantly lower in truck-weaned calves (adjusted means ± SE, 3.96 ± 0.01) compared to fence-line-weaned calves (4.03 ± 0.01) at post-weaning **(Sup Figure 2.4.B).** Likewise, a large proportion of the variation in the methanogenic gene portion of the rumen microbiome was attributed to weaning strategy and collection day (R^2= 8.1%, p<0.01, and R^2= 24.3%, p<0.01, respectively) **(Sup. Figure 2.4.C)**.

Differential abundance testing for post-weaning samples revealed that 64 methanogenic genera and 98 methanogenic genes had significantly lower relative abundance in fence-line-weaned calves, while 45 genera and 45 genes had significantly higher relative abundance, when compared to truck-weaned calves **(Figure 2.5).** Within the most abundant methanogenic genera and genes across samples we identified that the

archaeal genus *Candidatus Methanomethylophilus*, and some genes in the methylotrophic methanogenesis pathway (*mtmB* and *mtaB*) had significantly lower relative abundance in fence-line-weaned calves compared to truck-weaned calves; while archaeal genus *Methanobrevibacter* and some genes in the hydrogenotrophic methanogenesis pathway (*fwdB* and *fwdF*) had significantly higher relative abundance **(Sup. Table 2.5).**

Rumen resistome of beef calves was consistently dominated by tetracycline resistance genes

After trimming and removing host and dietary plant genomes, we assessed the Antimicrobial Resistance genes in the rumen microbiome using AMR++ v2, which uses a comprehensive database that includes 4 antimicrobial compound types [74]. The composition of antimicrobial resistance genes within the rumen (i.e., the rumen resistome) did not change significantly over time nor did it differ significantly between weaning groups. More than 90% of the resistome at the gene-group level was composed of tetracycline resistance genes, namely *Tet40, Tet44, TetO, TetQ* and *TetW*, with the latter comprising more than half of the total resistome content **(Figure 2.6.A)**. The homogeneous composition of tetracycline resistance genes was reflected in our analysis of both alpha and beta diversity, none of which demonstrated significant associations with weaning strategy (data not shown). In addition, differential abundance testing revealed limited statistically significant differences between weaning groups at weaning and post-weaning collection days **(Figure 2.6.B)**.

DISCUSSION

This study assessed the effect of castration timing and weaning strategy on the rumen microbiome of beef calves using a randomized controlled trial and longitudinal sampling. The dataset alone represents a substantial contribution to the limited body of literature pertaining to the rumen microbiome of beef calves specifically [18,55]. Using this dataset, we showed that the rumen microbiome shifted as calves approached weaning age, with a very noticeable and rapid change occurring within the first 48 hours after weaning. We did not find a significant effect of castration timing on the temporal dynamics of the rumen microbiome, although our sampling intervals may not have been frequent enough or close enough to the early castration events to capture short-term differences that may have occurred. Recent studies have found an association between the intestinal microbiome and increased adiposity [75] and growth inhibition [76], both of which can be impacted by the hormonal changes that occur with castration. Given this prior evidence and the limitations of our sampling design, we cannot definitively rule out an association between castration timing and rumen microbiome development, and thus more studies are warranted.

Our study showed that the changes in the rumen microbiome 48 hours post-weaning were significantly different in the calves weaned by fence-line compared to those weaned by truck. This effect was also observed when analyzing only the methanogenic genes and microbes of the rumen microbiome; specifically, *Methanobrevibacter* and some genes in the hydrogenotrophic methanogenic pathway were found in higher abundance in calves weaned by fence-line**,** while some genes in the methylotrophic methanogenesis pathway were in higher abundance in truck-weaned calves. In contrast, the antimicrobial resistance genes of the rumen microbiome (i.e., the rumen resistome) were not detectably altered by weaning, instead demonstrating a consistent dominance of tetracycline resistance genes across time and intervention groups.

Rumen microbiome dynamics around weaning could be driven by dietary and stress factors

The overall dominance of the phyla *Bacillota* (synonym Firmicutes) and *Bacteroidota* in the rumen microbiome of beef calves of weaning age was previously reported [19,22]. Both phyla seem to be dominant throughout rumen development in beef calves as previously reported [18]. The phyla *Actinomycetota* and *Fibrobacteres* had different relative abundances in previously reported rumen samples collected from beef calves at about the same weaning age; while the phyla *Verrucomicrobia* and *Tenericutes* represented less than 1% of total abundance in our study but were present in at least 2% relative abundance in other studies [19,22,55]. This heterogeneity in findings may be explained by several factors such as different genetics, diets and environments. In our study, the phyla *Bacillota* was highly abundant pre-weaning but decreased in relative abundance post-weaning. This phylum is represented mainly by the genera *Butyrivibrio* spp., a hemicellulose degrader and main producer of butyrate, which has been reported as highly abundant from birth to 96 days of age [18]. During this same three-week period just prior to weaning, the low abundant (<1%) phylum *Elusimicrobia* decreased remarkably across all calves. This phylum is an understudied anaerobic bacteria reported to be increased in feedlot finisher cattle [77] and in high-forage-fed dairy cattle during the dry period [78]. As previously described, the time-dependent changes of the rumen microbiome have been shown to be influenced by age and diet in cattle [15,19,68,79] and other ruminants [80,81].

Weaning strategy was associated with immediate differences in the rumen microbiome of the beef calves in this study. Weaning is an important physiological and life cycle event across all mammalian species; it can be especially important in livestock species because it involves not only dietary changes, but often concomitant separation from the dam, social regrouping, and transport to a new environment. The effects of weaning on animal productivity and welfare have been studied extensively [65,82,83], but their mechanisms and the overall benefits and drawbacks of different weaning strategies on various important phenotypes and production outcomes are still being investigated. The impact of weaning on the gastrointestinal microbiome has recently gained attention and has been described in swine [84], horses [85], sheep [86], goats [81] and dairy cattle [52], but to date not in beef cattle.

Consistent with other studies in cattle [52] and other livestock species [80,84], we observed a significant decrease in rumen microbial diversity shortly after weaning, which was especially marked in truck-weaned calves. Stressors such as heat [87] and long-distance ground transportation [88] have also been associated with decreased rumen microbiome diversity in cattle, suggesting that physiological stressors at the host level may also manifest as decreased diversity within host-associated microbiomes. Interestingly, the rumen of truck-weaned calves had the lowest microbial diversity across all collection days and groups, which may be explained by the additional stressors these calves experienced due to ground transportation, physical separation from their dams, and change of environment as well as diet. The mechanisms involved in the response of gastrointestinal microbiomes to physiological stressors are still unclear but may include: oxidative stress, erratic activation of immune response against bacteria and the secretion of bacterial toxins [89–91]. To note, our study did not measure stress levels using biological markers and thus did not present our results in terms of “high” or “low” stress weaning methods. However, the body of evidence shows that sequential weaning (i.e., via fence line) is associated with decreased levels of stress-associated biomarkers compared to abrupt weaning by long-distance transportation [65,82].

We observed a significant increase in the relative abundance of the 2 most abundant genera, *Fibrobacter* spp. and *Prevotella* spp., in the 48 hours after weaning. *Fibrobacter* spp. are cellulose degraders previously identified as part of the core heritable rumen microbiome [9] that colonizes the rumen of beef calves after 7 days of age [18]. The transition of calves to a forage- or grass-exclusive diet may provide a fiber-rich substrate that supported the relative increase of this genus with the rumen microbial community. The increase of *Prevotella* after weaning and during dietary transitions has been described in cattle [15,52,79], pigs [84] and horses [85]. This genus is composed of several species with a variety of biological functions, including use of readily available carbohydrates, degradation of hemicellulose, and protein and peptide breakdown [92,93]. The wide metabolic plasticity of the genus *Prevotella* offers evidence of the complex functional profile of the rumen microbiome. For instance, some *Prevotella* species have been associated simultaneously with both low and high methane emission cattle [32], while other species such as *Prevotella ruminicola* have been identified as the

bacterial host of antimicrobial resistance genes in the cattle rumen [40]. Interestingly, increases in the abundance of this genus in the rumen have also been associated with other non-dietary stressors in cattle [56,57], suggesting a role of microbes in the physiological response of cattle to diverse stressors. It is challenging to identify a single or definitive mechanistic link between rumen microbiome variation and host-level stressors such as a change of diet and weaning. In humans, the gut microbiome is considered part of the gut-brain axis, which helps regulate stress through processes that include vagus nerve modulation, gut hormone signaling, the immune system and microbial metabolites [94]. The role of rumen microbes in the stress response of ruminants is still unclear and requires further study.

Weaning as a process represents numerous changes (i.e., diet, age and change of environment), all of which can impact both the bovid host and its associated microbes. Our study was not designed to isolate the effect of each factor, but instead to describe the total effect of typical weaning processes as a multi-factor event that occurs during the cattle life cycle. Dietary shifts and aging are co-occurring factors that affect the rumen microbiome and are nearly impossible to disentangle. Nonetheless, we tried to limit the potentially confounding effect of other covariates such as pre-weaning diet and genetic background, while also appropriately mimicking two common weaning practices in the US [83], as they happen in the field. Our study did not include long-term sampling to understand the long-term effects of weaning on the rumen microbiome. If they exist, such long-lasting shifts in the rumen microbiome may be highly relevant to important animal health and production outcomes, and thus should be addressed by future studies.

Ruminal methanogenic genes may be more readily influenced by external factors than antimicrobial resistance genes

The cattle rumen produces methane [95], and the archaeal community (particularly the phylum *Euryarchaeota*) contains many of the main methane-associated microbes. These microbial taxa are found in the rumen microbiome from an early age and may be part of the initial microbial colonization of the rumen [17,18]. Through a variety of studies, a set of non-archaeal microbes have also been associated to high methane emission in cattle: *Methanobrevibacter (particularly Mbb. Gottschalkii and Mbb. Ruminantium),*

Christensenellaceae, Mogibacteriaceae, Ruminococcaceae, Lachnospiraceae and Rikenellaceae [24,34,96,97]. On the other hand, several others are highly abundant in low methane emitters: *Methanosphaera, Vellionellales* and *Desulfovibrionales* [24,98]. The literature in this area is still somewhat ambiguous, but previous findings combined with environmental evidence have allowed the development of a preliminary database of methane-associated bacteria, archaea and genes [73]. Using this tool, we observed a significant association between weaning strategy and the composition of methane-associated genes in the rumen of beef calves. From 10 different methane cycle pathways in the database [73], rumen methanogens only use 3: hydrogenotrophic, methylotrophic and acetoclastic [7,95]. We identified the genus *Methanobrevibacter* and genes from the hydrogenotrophic (*fwdB* and *fwdF*) methanogenic pathways in higher relative abundance in fence-line-weaned calves compared to truck-weaned calves. Considering that *Methanobrevibacter* is an important methanogenic archaeon and that the hydrogenotrophic pathway is the main methanogenic pathway in ruminants [95], our findings may suggest that the ruminal microbial community of calves weaned by fence line could have an increased capacity to generate methane, at least in the very immediate post-weaning period.

However, we caution against over-interpretation of these results, particularly because our analysis is conducted at the DNA level, which does not necessarily correlate with transcription and production of metabolites such as methane. Proper correlation with metatranscriptomics, proteomics, metabolomics and/or phenotypic testing remains to be elucidated in future studies. Additionally, more extensive phenotypic measurements would provide more actionable results, including respiratory chambers or antimicrobial susceptibility testing to measure methane and phenotypic antimicrobial resistance, respectively. Since these metrics are beyond the scope of this study, the reader should not extrapolate our DNA-level results to phenotypic expression. Moreover, metagenomic studies require replication, particularly given the relatively sparse information and sequence databases available for methane-production pathways and microbes. Specifically, the lack of gene ontology annotation in the available methanogenic database prevented an aggregated pathway-level analysis, and thus we were only able to analyze the rumen microbiome on a gene-by-gene basis. Given that methane production is typically the result of complex gene-gene interactions, the results

generated from this gene-by-gene approach need further validation. A rumen-specific methanogenic database with comprehensive annotations is a critical gap for future metagenomic studies of the rumen microbiome and methane production.

Recent evidence suggests that the rumen can be a potential source of antimicrobial resistance genes, with a highly diverse and concentrated microbial community that can favor horizontal gene transfer [30,99,100]. Our study found a very consistent and dominant distribution of tetracycline (*tet40, tet44, tetO, tetQ, tetW*) and nitroimidazole (*nimJ*) resistance genes within the rumen resistome. Our findings regarding the dominance of tetracycline resistance genes are consistent with previous studies of the rumen resistome of dairy cattle, both at the DNA [101,102] and RNA levels [100]. Furthermore, this pattern of a tetracycline-dominated resistome has been described in numerous beef and dairy resistome studies investigating different sample types (e.g., feces, soil and water) and even different countries [103–105]. Contrasting to our results, chloramphenicol, microcin, aminoglycoside and streptomycin resistance genes have been reported to be more prevalent in the rumen of adult beef cattle not exposed to antibiotics [30]. While concentrate-based diet [30] and even milking traits [102] have been associated with differences in the rumen resistome, we did not observe an association with age, castration timing or weaning strategy in beef calves. Interestingly, the predominant tetracycline resistance pattern in the rumen is reflected in feces, as described not only in dairy calves [106] but also in unexposed wild ungulate species (elk and bison). The presence of tetracycline resistance genes in wild ruminants suggests that this phenomenon may have broader origin in wild animals [107].

The most abundant gene across all samples in this study was *tetW*. Recently, a high abundance of tetW transcription within the rumen of beef cattle was reported using a metatranscriptomics approach; and the carbohydrate degraders *Ruminococcus* spp., *Prevotella ruminicola*, *Muribaculaceae* spp. and *Collinsella aerofaciens* were listed as common bacterial hosts of expressed ARGs [40]. Additionally, the highly abundant *tetW* gene has been found located in a novel integrative and conjugative element in the ruminal community [100], supporting the hypothesis that horizontal gene transfer of AMR genes within the rich and complex microbial community of the rumen supports the abundance and persistence of tetW in the rumen. Further research is needed to both

replicate this finding and to understand its importance, considering the rumen microbiome of cattle not only as a potential source of antimicrobial resistance genes [30], but also as a potential ecosystem favorable to increased horizontal gene transfer [95].

Microbial dark matter dominated the rumen metagenomic data

The use of metagenomics in microbiome research has drawn attention to the high amount of still unidentified genomic material that makes up these communities [108]. The proportion of unclassified sequences, referred to as microbial dark matter [108] or dark microbiome [109], varies depending on the niche. Between 25-81% [108,110] of microbial genomic material in diverse environmental niches has been cataloged as unknown. Whereas in potentially less diverse niches the unclassified sequences are lower. For instance, around 50% of non-host sequences have been reported as unclassified in the microbiome of *Arabidopsis thaliana* leaves [111], while between 2-4% of sequences in industrial food ingredients could not be identified [112]. Although well documented, this limitation is scarcely reported and addressed in rumen microbiome research. From 22 rumen microbiome studies that cited kraken [113] or kraken2 [114] as taxonomic classifier in Pubmed (accessed on April 2023), only 5 addressed the classification rate issue [13,45–48].

Three of the 5 studies showed that the use of the Hungate project genomes and [72] their own (self-produced) Metagenome-assembled genomes (MAGs) to customize the reference database for the taxonomic classification of rumen microbiome, increased the classification rate on average up to 50% [47], 62.6% [45] and 70% [48] on their studies. Despite the substantial increase in classification rate with MAGs, a recent study evidenced that incomplete or informal taxonomic lineages of MAGs (i.e., lack of appropriate labels at every taxonomic rank) limit the classification at lower taxonomic levels (i.e., genus or species) [46]. Interestingly, one of these 5 studies identified on average 12% of the previously unclassified metagenomic reads as ciliates by adding 52 high-quality rumen ciliate genomes to their reference database [13]. Altogether, these efforts highlighted the importance of reference database customization with rumen-specific organisms (bacterial and non-bacterial) to increase the classification rate and revealed how poorly explored is the rumen microbiome. For this reason, we increased

our classification rate while also ensured an accurate genus-level classification by customizing our reference database with the genomes of the Hungate project and the most comprehensive collection of archaea, bacteria, virus, plasmids, human, UniVec_core, protozoa and fungi from RefSeq NCBI.

There is a serious bias in reference databases that limit our interpretation not only of the taxonomic profile of microbial communities but also of the functional characterization of this microbial dark matter [115]. When it comes to the rumen microbiome, said limitations can hinder numerous potential biotechnological applications of their unknown microorganisms. As potential solutions, some research efforts have added up to the hungate project [72], exploring the yet-unknown microorganisms of the rumen through culturomics [116], and even more, investigating the neglected viral [117] and plasmid [118] communities of the rumen. More studies are warranted to shed some light on this rumen microbial dark matter.

CONCLUSION

The rumen microbiome of beef calves is a complex and dynamic community that shift around weaning. Significant changes in the taxonomic and functional profile of the rumen microbiome of beef calves are associated with weaning strategy within the first 48 hours after weaning. Unlike methanogenic genes, the rumen resistome was not impacted by time neither weaning strategy. Castration timing was not associated to rumen microbiome changes in this study. However, given the sampling design constraints an association cannot be definitively ruled out. More studies are warranted to describe the short-term effects of castration timing and long-term effects of weaning strategy on the rumen microbiome, while overcoming the methodological limitations of shotgun metagenomics to investigate this unique microbial community.

METHODS

We conducted a randomized longitudinal "2x4" factorial study with 32 beef calves randomly assigned to 2 different weaning strategies nested within 4 different castration timing windows. Rumen fluid samples were collected at pre-weaning, at weaning and at post-weaning. DNA was extracted and submitted for shotgun metagenomic sequencing. Bioinformatic analysis was used to determine the composition of the rumen microbiome for taxa features (taxonomic profiling) and gene features (functional profiling) and was represented in terms of: (1) feature relative abundance, (2) alpha diversity (richness, Shannon index and Pielou's index), (3) beta diversity and (4) feature differential abundance testing. Differences in these metrics were compared between collection days, castration timing and weaning strategy groups.

Study design and interventions

This randomized controlled trial was conducted in a single cow-calf herd at the North Central Research and Outreach Center (NCROC) at the University of Minnesota (Grand Rapids, MN) from March to October 2021. The herd contained around 120 certified Angus cows raised on 801,278 m^2 of mixed pasture. Our study assessed two management interventions in a 4x2 factorial design: 4 different castration timings and 2 weaning strategies. All bull calves born in the 2021 season were eligible for study enrollment unless they were born under dystocia conditions, with a visible abnormality or disease. Thirty-two bull calves were enrolled at birth consecutively in a 26-day window of time (March – April). Animals were randomized at birth to both interventions with a balanced and crossed design, with 8 animals per each castration timing group and 16 animals per weaning strategy group. The study calves (N=32) had similar genetic backgrounds, were fed the same diet, were turned out in a single group and, except for castration timing, managed under the same standard procedures until weaning.

The 4 castration timing groups assessed were: castration within 48 hours of birth (N=8, March 14th – April 9th), at turn-out (N=8, May 25th), at pre-weaning (N=8, September 21st) and at weaning (N=8, October 18th). Birth and turnout castrations were done using the Ideal® Calf and Lamb Bander (Neogen, USA). Briefly, the testicles and scrotum were

pulled down, the band was opened and placed up over the scrotum, and after checking that the testicles were still in the scrotum, the band was released just above the top of the testicles. A final check was done to ensure both testicles were still in the tip of the scrotum and that the ring was placed properly. Pre-weaning and weaning castrations were done in the chute located at the handling facility following a similar procedure but using a XL Castrating Bander (Wadsworth Manufacturing, USA). All animals, regardless of their castration group, were checked for testicles at pre-weaning and weaning collection days. As part of the health management plan, several treatments and vaccinations were administered to the study calves on the same days as the castrations, as follows: at turn-out Ultrachoice® 8 (Zoetis, USA), Inforce® 3 (Zoetis, USA), Nuplura® PH (Elanco, USA) and Cydectin® (Elanco, USA); at pre-weaning Titanium® 5 (Elanco, USA), Nuplura® PH (Elanco, USA) and Ultrachoice® 8 (Zoetis, USA); at weaning Valbazen® (Zoetis, USA), Titanium® 5 (Elanco, USA), Nuplura® PH (Elanco, USA) and Ultrachoice® 8 (Zoetis, USA).

The 2 weaning strategies assessed were weaning by fence-line and truck. On the day of weaning, all study calves (N=32) were brought through the chute for sample collection, and the calves in the "at-weaning" castration group were banded (N=8). Upon exiting the chute, calves assigned to the "fence-line weaning" strategy (N=16) were brought to a pasture that adjoined a pasture housing their dams, but with separation via an electric fence; both pastures contained a mix of Kentucky bluegrass (*Poa pratensis*), tall fescues (*Festuca arundinacea*), red clover (*Trifolium pratense*), timothy grass (*Phleum pratense*), orchard grass (*Dactylis glomerata*), perennial ryegrass (*Lolium perenne*), smooth brome (*Bromus inermis*), with no extra dietary supplementation. Calves assigned to the "truck weaning" strategy (N=16) exited the chute and were assembled in a preloading pen and then loaded onto a truck and transported for 2 hours. After transport, they were unloaded into a feedlot-sized pen at the south station of NCROC, where they were kept in a roofed pen with a thick layer of straw bedding and a J-bunk concrete feeder. They were given a moderate quality 25/75 alfalfa to grass mix, supplemented by Wind and Rain® mineral and American Stockman® salt. This formulation was intended to better reflect the pasture-based diet of the fence-line weaning group, i.e., a grass-based diet without concentrate-based supplementation.

Ruminal samples were collected on 3 collection days: at pre-weaning (September 21st); at weaning (October 18th); and at 48 hours post-weaning (October 20th). Weights were collected on these days and at birth and turn-out. For each collection day, all study calves were run through the chute to collect ruminal samples. Calves assigned to the pre-weaning castration group were banded in the chute at pre-weaning collection day, and calves assigned to the weaning castration group were banded in the chute at weaning collection day. Thus, at pre-weaning collection, half the calves had already been castrated (8 at birth and 8 at turnout); at the weaning collection, 24 of the calves had already been castrated (8 at birth, 8 at turnout and 8 at pre-weaning); and at post-weaning collection, all calves had already been castrated. Thus, the castration groups varied depending on the collection day: 3 castration strategies were contrasted at pre-weaning (Birth, Turn-out and Not castrated); 4 at weaning (Birth, Turn-out, Pre-weaning and Not castrated); and 4 at post-weaning (Birth, Turn-out, Pre-weaning and Weaning). At the same time, 2 weaning strategies were assessed: Fence-line and Truck **(Supp. Figure 1).**

Sample collection and DNA extraction

Sample collection and animal handling were done following ethical guidelines approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Minnesota. Rumen fluid was collected by esophageal tubing using a Frick's speculum adapted to a collection flask and vacuum pump. Between each animal, the tubing and collection flask were emptied, disinfected with sodium hypochlorite at approximately 10% concentration, and rinsed thoroughly with tap water. The tube was inserted into the oral cavity and advanced down the esophagus until the fiber mat was reached, at which point the tube was retracted 5-8 cm in order to obtain fluid. Rumen fluid was collected in 50 ml sterile tubes, immediately transported to the laboratory at approximately 4°C and stored at -80°C within 4 hours of collection. Weight was collected from each calf using a floor scale adapted to the cattle chute. The average daily gain (ADG) was obtained individually for each animal, subtracting the birth weight to the post-weaning weight and dividing the result by the age in days. The cohort of 32 calves was monitored closely by NCROC staff, who reported any variation in diet, behavior, or health issues.

DNA was extracted from each sample in randomized batches of 12 samples under aseptic conditions to avoid batch effect and cross-contamination. DNA extraction blanks consisting of CD1 buffer were used as negative controls. To begin DNA extraction, rumen fluid was thawed and homogenized by vortex for 3 minutes. An aliquot of 1 ml was centrifuged at 16 000 rcf for 10 minutes in an Eppendorf 5415R centrifuge at room temperature. Supernatant was discarded, and the remaining pellet was used as initial sample for column-based DNA extraction using the Dneasy® PowerSoil® Pro Kit (QIAGEN, USA) following the manufacturer's protocol without modifications. Briefly, the pellet was resuspended in 800 μl lysis buffer (CD1 solution), suspension was transferred to PowerBead Pro® Tubes (zirconium beads), then bead beating was performed in 3 cycles of 20 seconds at 2,200 rpm with 30 seconds intervals and centrifuged at 16 000 rpm for 2 minutes. Finally, 600 μl supernatant were transferred to QIAcube Connect® equipment (QIAGEN, USA) for a fully automated DNA extraction.

Library preparation and shotgun metagenomic sequencing

DNA extractions were submitted to the University of Minnesota Genomics Center (UMGC). Along with sample DNA and negative controls, mock community DNA (ZymoBIOMICS Microbial Community Standard II Log Distribution – Catalog N° 6310) already extracted following the process described above, was included as a positive control for the library preparation and sequencing process. DNA quantity and quality was assessed using the PicoGreen assay (Thermo Fisher, USA) and 260/230 ratio in Nanodrop1000 (Thermo Fisher, USA), respectively. Barcoded libraries were generated using Illumina Nextera XT DNA library preparation kit (Illumina, USA) following manufacturer's protocol. Shotgun metagenomic paired-end sequencing (2x150 bp) was performed in a single pool across 2 lanes of S4 flow cells (2,250 million reads/lane expected) of a NovaSeq 600 platform (Illumina, USA) using kit v1.5 (300 cycles).

Bioinformatic analysis

Demultiplexed paired-end sequencing reads were analyzed using the AMR ++ version 2.0 pipeline [74]. This suite includes quality-based trimming and filtering of sequencing reads using Trimmommatic [119], alignment of surviving high-quality reads to the host genome using Burrows-Wheeler-Aligner (BWA) [120], removal of host-aligned reads by

BEDTools [121], and resistome and microbiome taxonomic profiling using the filtered non-host reads as input.

The initial protocol for microbiome taxonomic profiling included the following standard processes: (1) host-decontamination aligning high-quality reads to *Bos taurus* reference genome (Genome Bos_taurus_UMD_3.1, accession number: GCA_002263795.3), (2) taxonomic classification with kraken2 [114] using a confidence score of 0 and the standard genome database (accessed in July 2020), which contained archaea, bacteria, virus, plasmids, UniVec_Core and the human genome. As this protocol was not able to classify a high proportion of sequencing reads, we attempted a classification rate increase by a customized rumen-specific protocol detailed as follows: (1) For the host-reads decontamination step we aligned the high-quality reads to *Bos taurus* reference genome (Genome Bos_taurus_UMD_3.1, accession number: GCA_002263795.3) and dietary plants genomes available in GenBank, which included: *Trifolium pratense* (full genome, accession: ARS_RC_1.1), *Dactylis glomerata* (full genome, accession: GCA_007115705.1), *Lolium perenne* (full genome, accession: MPB_Lper_Kyuss_1697), *Poa pratensis* (chloroplast genome, accession: NC_057962.1), *Festuca arundinacea* (chloroplast genome, accession: NC_011713.2), *Phleum pratense* (chloroplast genome, accession: NC_067044.1), *Bromus inermis* (chloroplast genome, accession: NC_067047.1), and *Festuca pratensis* (plastid genome, accession: NC_019650.1); (2) taxonomic profiling using Kraken2, with a confidence parameter at 0.1 to decrease the likelihood of spurious (i.e., false positive) classifications [122]. Instead of using the default kraken2 database, we built a custom database that included reference genomes from RefSeq's NCBI for archaea, bacteria, virus, plasmids, human, UniVec_core, protozoa and fungi (accessed in January 2023). In addition, we added genomes obtained from 410 rumen-specific bacteria isolated for the Hungate project [72] (source: https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=HungateCollection). It was previously shown that including these genomes significantly increases the taxonomic classification of rumen microbiome samples [46].
The resistome analysis, implemented in AMR++, was carried out by aligning filtered non-host reads to the MEGARes 2.0 database [74] using BWA [120]. Antimicrobial resistance genes (ARGs) that obtained at least 80% gene fraction were considered present in the respective sample, and all aligning reads were output to a count matrix

used for downstream analysis. ARGs labeled as '*RequiresSNPConfirmation'* were excluded from the count matrix and not considered for downstream analysis.

The taxonomic and functional profiling for methanogenic microbes and associated genes was done using McycDB database and scripts, using default pipeline and tool settings [73]. Briefly, for functional profiling, the forward and reverse non-host reads were merged using PEAR [123] and then a translated search in the McycDB database was done using DIAMOND [124] with e-value= 1e-4. For methanogenic analysis only, the number of sequences in every sample was normalized by random subsampling to 6,188,129 reads per sample (the minimum number of sequences per sample within our dataset) for downstream analysis. Results were reported at the gene level. For taxonomic profiling of methanogenic taxa, in addition to the previously described steps, the merged reads matching a methanogenic gene were subsetted using SeqKit [125] and taxonomically classified using Kraken2 [122]. Results were reported at the phylum, class, order, family, genus and species levels.

Feature and OTU data analysis

The feature or Operational Taxonomic Unit (OTU) count matrices obtained by the bioinformatic analyses, and the study metadata, were used for descriptive microbiome analysis in R (version 4.1.0, https://www.r-project.org/). We analyzed the relative abundances of features and OTUs; alpha diversity (richness, Shannon's and Pielou's indices); beta diversity (Bray-Curtis distance method); and performed differential abundance testing with the *phyloseq* [126], *metagenomeseq* [127] and *vegan* [128] packages. The first metric represents the percentage of total abundance that one feature represents in the microbiome (usually in the 0-100% scale); the second metric represents the variance within each sample considering number, distribution and abundance of features; the third one represents the variance between samples expressed as a distance matrix; the fourth is used to identify specific features that differ between 2 groups of samples [129]. Data visualization was performed with the *ggplot2* package [130].

Statistical analysis

Statistical analysis was done in R using the *metagenomeseq [127]*, *limma [131], vegan [128]*, *lme4 [132]*, *lmerTest [133]*, *emmeans [134]* and *car [135]* packages. The major independent (predictor) variables were the 2 interventions (castration timing and weaning strategy), collection day and their interaction when appropriate. Unless otherwise stated, statistical significance was considered at a P-value < 0.05.

For alpha diversity, the dependent (outcome) variables were richness and Shannon's indices calculated at the genus level. We used a linear mixed-effect model considering animal ID as a random effect (to account for repeated measures over time on each calf) and independent variables (castration timing, weaning strategy and their interaction) as fixed effects. Type-III ANOVA was used to assess statistical significance for each independent variable in the model. When statistically significant, a post-hoc comparison was done between groups with least square means. For differential abundance testing, CSS-normalized OTU counts at the phylum and genus levels were considered as dependent variables. We used multivariate zero-inflated Gaussian mixture models with animal ID as a random effect and independent variables as fixed effects. When statistically significant, pairwise comparison of the log2-fold change between groups was performed with Benjamin-Hochberg (BH) p-value correction for multiple comparisons.

For beta diversity, the dependent variable was dissimilarity index calculated at the genus level using Bray-Curtis distances. We performed PERMANOVA to infer statistically significant associations between dependent variables and beta diversity ordination. The R2 value was used to estimate the amount of variation partitioned to a given dependent variable and the associated p-value to determine statistical significance. We used Kruskal-Wallis or ANOVA to assess statistical significance of other independent but potentially confounding variables (e.g., number of raw sequencing reads, number of classified reads, average daily gain and age) to determine whether to include them as potential covariates in the regression models.

TABLES & FIGURES

Table 2.1. Summary of calf ages and weights, by castration timing and weaning groups (mean ± SE)

Variable	Castration				p-value*	Weaning		p-value*
	Birth	Turnout	Pre-weaning	Weaning		Fence-line	Truck	
Age at Post-weaning (days)	213 ± 2.56	212 ± 1.56	209 ± 3.32	209 ± 3.71	0.8	**208 [a] ±2.17**	**214 [b] ±1.63**	**0.01**
Weight at birth (kg)	38.21 ± 1.99	37.06 ± 1.74	38.1 ± 1.66	37.64 ± 1.07	0.9	38.52 ± 0.95	37.07 ± 1.24	0.2
Weight at Post-weaning (kg)	263.83 ± 7.67	267.44 ± 8.17	267.18 ±11.08	275.06 ± 8.95	0.8	276.52 ± 6.89	260.8 ± 5.05	0.1
Average daily gain (kg)	1.06 ± 0.04	1.09 ± 0.04	1.1 ± 0.05	1.14 ± 0.05	0.6	**1.15 [c] ± 0.03**	**1.04 [d] ± 0.02**	**0.02**

***Kruskal-Wallis test, values with different superscript letters were significantly different**

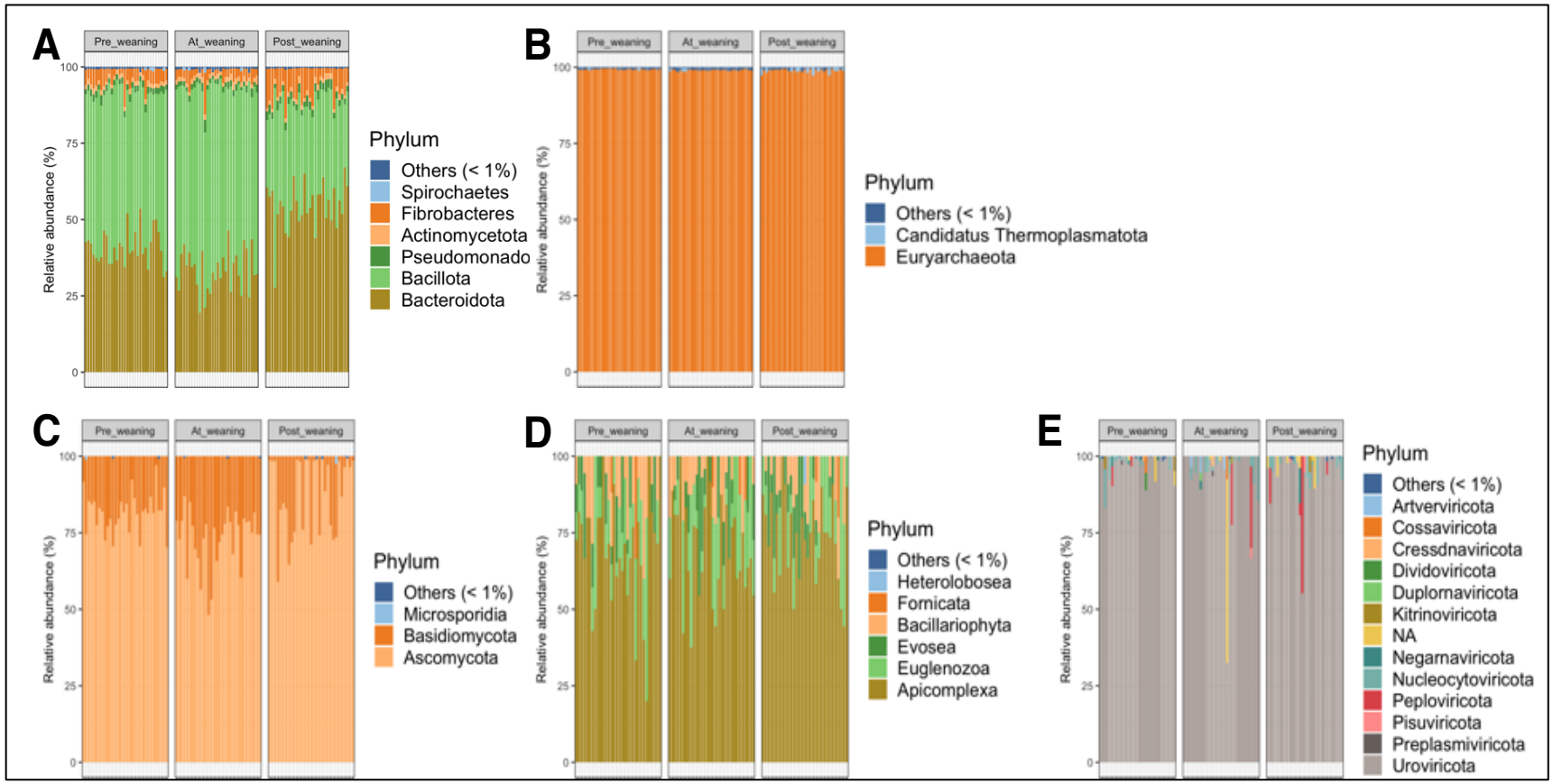


Figure 2.1. Relative abundance plots of phylum-level microbiome composition for A) Bacteria, B) Archaea, C) Fungi, D) Protozoa, E) Virus, grouped by collection day (Pre_weaning, At_weaning and Post_weaning). Phyla with < 1% relative abundance are grouped as "Others". Each bar represents one sample.

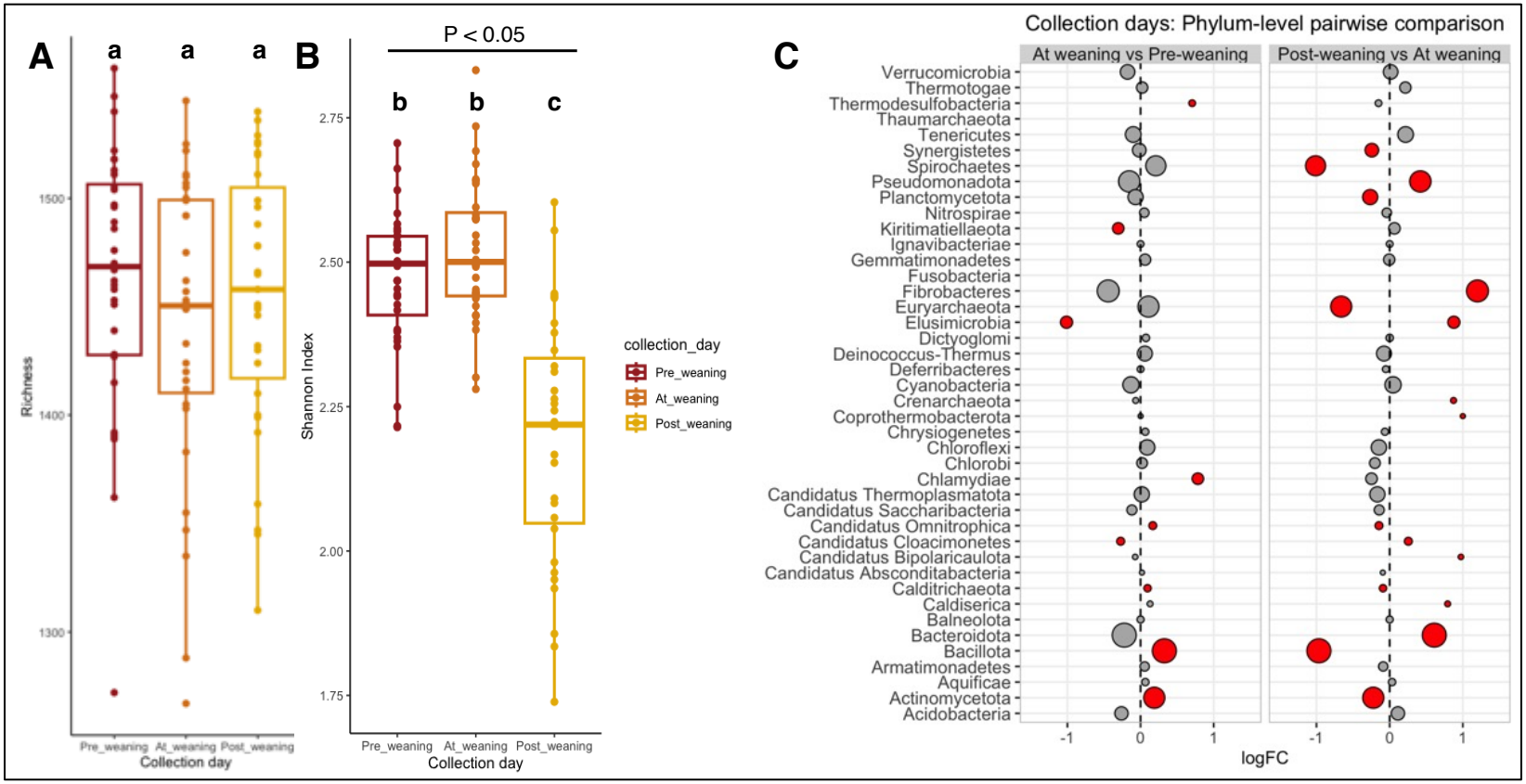


Figure 2.2. Rumen microbiome, comparisons over time. Box plots of A) richness and B) Shannon's index, at the genus level, grouped by collection day. Boxes represent the 25th to 75th percentile; horizontal line represents the median; and whiskers indicate 1.5× the interquartile range (IQR), *P*-values from Type-III ANOVA, collection days with different superscript letters were significantly different. **C) Differential abundance of phylum-level counts between collection days, expressed as log_2 fold change (LogFC).** Statistically significant logFC values (adjusted *P*<0.05) are depicted in red, and non-significant in grey. Circle diameter is proportional to the average abundance of each phylum across all samples.

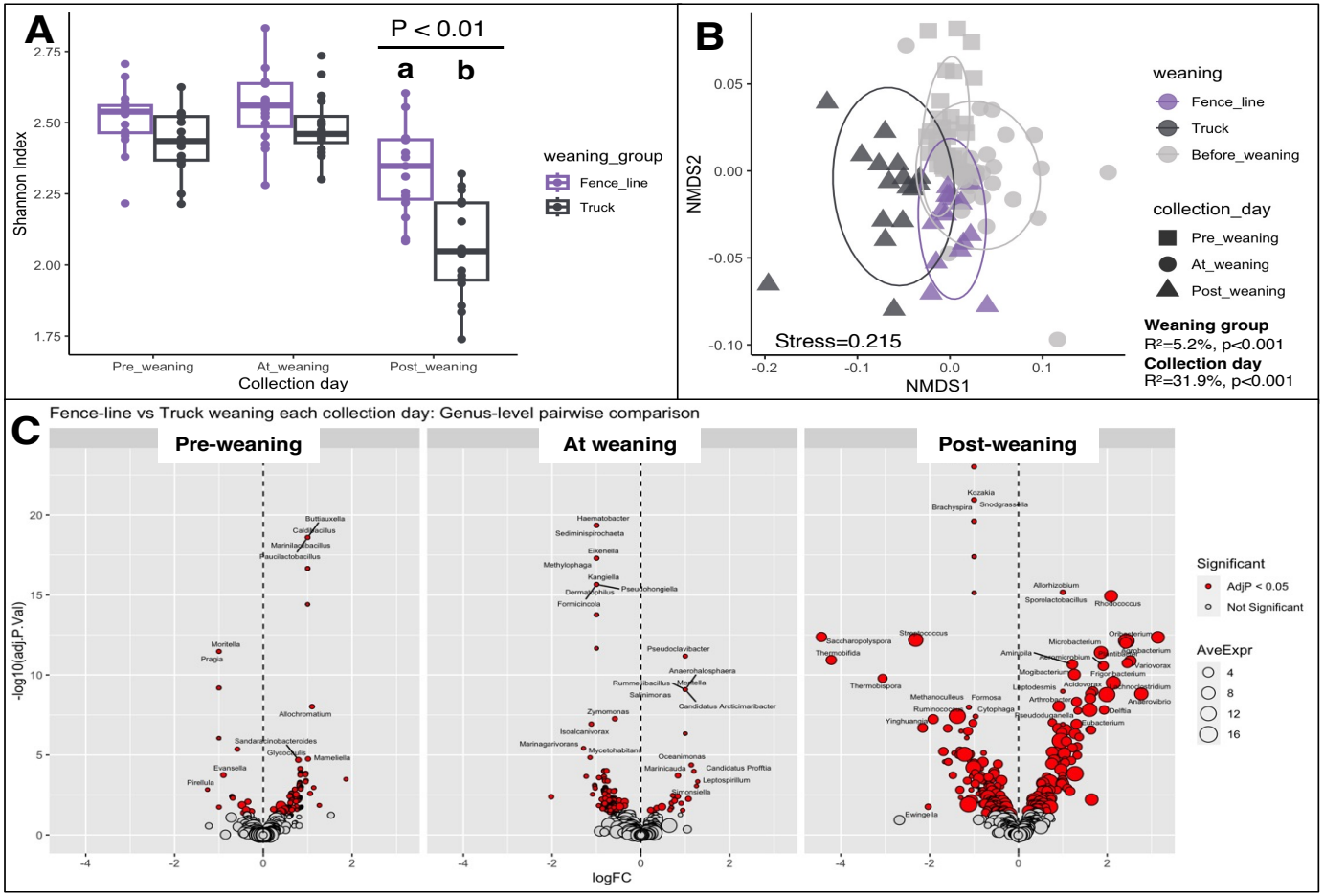


Figure 2.3. Rumen microbiome, differences by weaning strategy. A) Box plots of Shannon's Index at the genus level, stratified by weaning strategy across collection days. Boxes represent the 25th to 75th percentile; horizontal line represents the median; and whiskers indicate 1.5× the interquartile range (IQR), *P*-values from Type-III ANOVA, weaning groups with different superscript letters were significantly different. **B) Non-metric multidimensional scaling (NMDS) ordination plots based on Bray–Curtis distances at the genus level colored by weaning strategy.** *P*-value and R2 values from PERMANOVA testing. **C) Differential abundance of microbial genera between fence-line and truck weaned calves at different collection days, expressed as log2 fold change (LogFC).** Statistically significant logFC (adjusted *P*<0.05) are depicted in red and non-significant in grey. Circle diameter is proportional to the average abundance of each phylum across all samples.

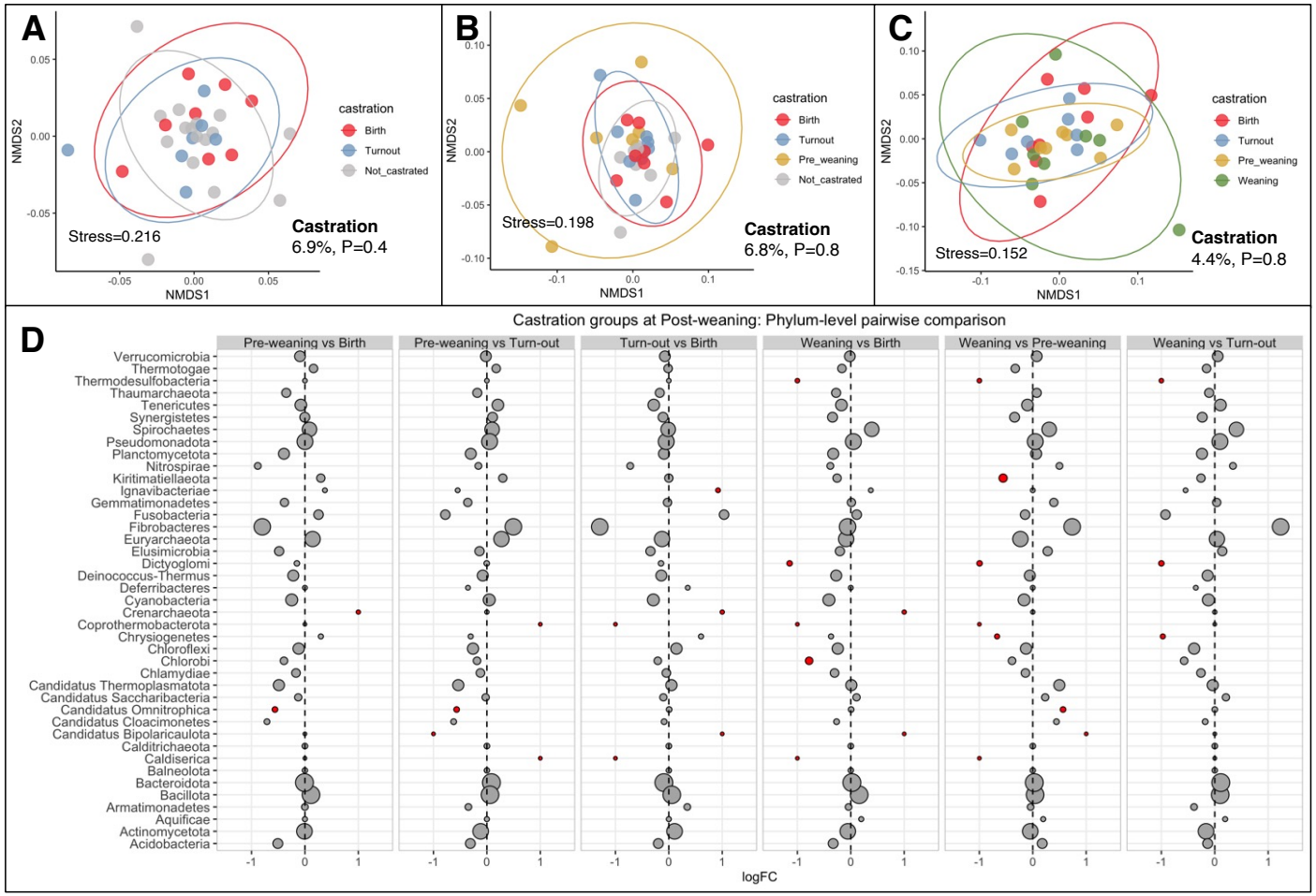


Figure 2.4. Rumen microbiome, differences by castration timing. Non-metric multidimensional scaling (NMDS) ordination plots based on Bray–Curtis distances at the genus level for A) Pre-weaning, B) At weaning, and C) Post-weaning collection days, colored by castration timing group. *P*-value and R2 values from PERMANOVA testing **D) Differential abundance of microbial phyla between castration timing groups at post-weaning day expressed as log2 fold change (LogFC).** Statically significant logFC (adjusted *P*<0.05) are depicted in red and non-significant in grey. Circle diameter is proportional to the average abundance of each phylum across all samples.

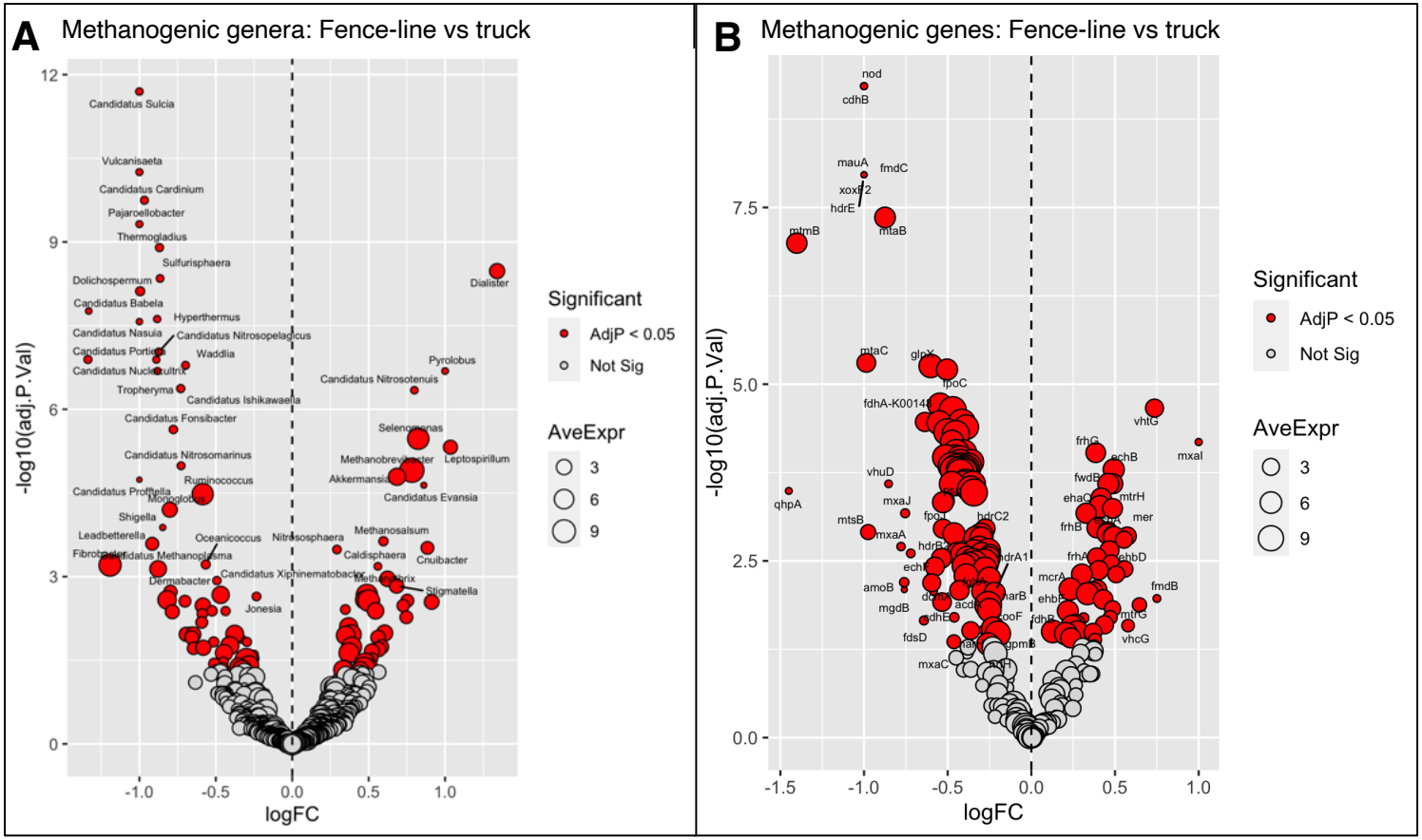


Figure 2.5. Differential abundance of methanogenic microbes (A) and genes (B) within rumen microbiome samples collected at 48 hours post-weaning, compared between weaning strategy and expressed as log2 fold change (LogFC). Statistically significant log2FC (Benjamin-Hochberg adjusted P<0.05) are depicted in red, non-significant are depicted in grey. Circle diameter is proportional to the average abundance of each feature (i.e., genus or gene) across all samples.

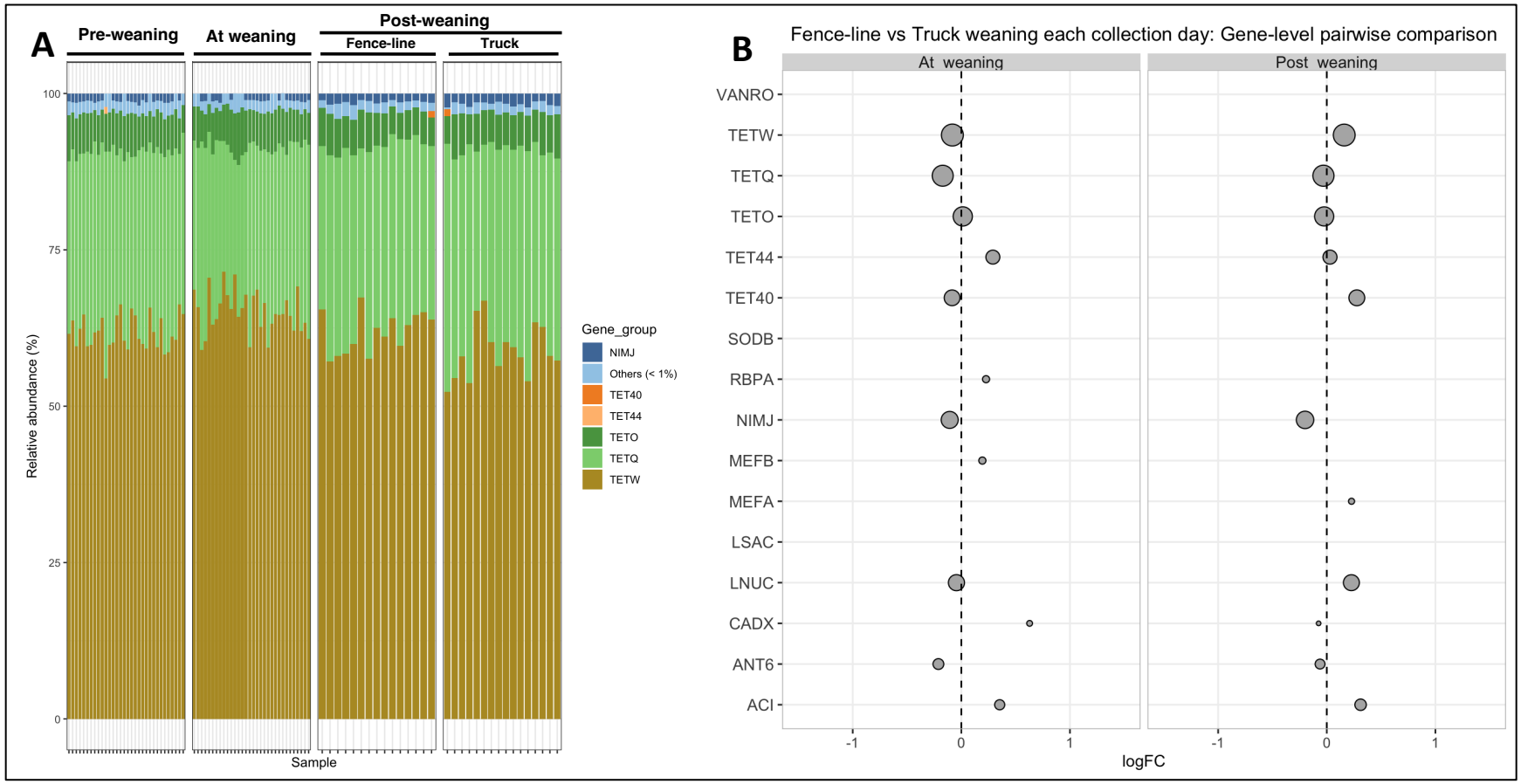


Figure 2.6. Rumen resistome by weaning strategy across collection days. A) Relative abundance plot of antimicrobial resistance genes (ARGs) at the ARG group level, grouped by collection day and weaning strategy. ARGs with < 1% abundance are grouped as "Others". Each bar corresponds to an individual sample. **B) Differential abundance of ARG groups between fence-line and truck weaned calves, at weaning and post-weaning days, expressed as log2 fold change (LogFC).** Statistically significant logFC (adjusted P<0.05) are depicted in red and non-significant in grey. Circle diameter is proportional to the average abundance of each ARG group across all samples.

SUPPLEMENTARY MATERIAL

Supplementary Table 2.1. Number of reads per sample analyzed by 2 taxonomic profiling protocols (mean [min – max])

Step	Standard protocol			Customized protocol		
	Sequencing reads per sample (x10^6)	Proportion of initial sequencing reads (%)	Proportion of non-host reads (%)	Sequencing reads per sample (x10^6)	Proportion of initial sequencing reads (%)	Proportion of non-host reads (%)
Raw sequences	54.2 [30.5 - 75.8]	100	-	54.2 [30.5 - 75.8]	100	-
Sequence trimming	52.3 [29.2 - 73.2]	96.5	-	52.3 [29.2 - 73.2]	96.5	-
Host decontamination	41.9 [24.9 - 59.9]	77.3	100	38.3 [23.3 - 55.2]	70.7	100
Taxonomic classification						
Confidence score = 0	5.9 [3.5 - 8.2]	10.9	14.1	8 [4.9 - 11.3]	14.8	21
Confidence score = 0.1	1.3 [0.7 - 2]	2.4	3.1	1.6 [0.9 - 2.3]	3	4.2

Supplementary Table 2.2. Read classification rates by taxonomic level.

Taxonomic Rank	Number of rank-specific classified reads/non-host reads (%)	Number of rank-specific classified reads/total number of classified reads (%)
Domain	4.15	100.0
Kingdom	3.99	96.2
Phylum	3.99	96.2
Class	3.95	95.1
Order	3.94	94.9
Family	3.87	93.2
Genus	3.77	91.0
Species	3.14	75.7

Supplementary table 2.3. Number of unique OTUs identified at different taxonomic levels

Domain/sub-kingdom	% of total classified reads*	Phylum	Class	Order	Family	Genus	Species
Archaea	1.365	7	18	31	48	140	351
Bacteria	98.589	40	93	219	519	1909	8674
Virus	0.006	14	25	34	77	368	390
Protozoa	0.001	6	11	13	15	18	40
Fungi	0.017	3	12	16	27	48	83

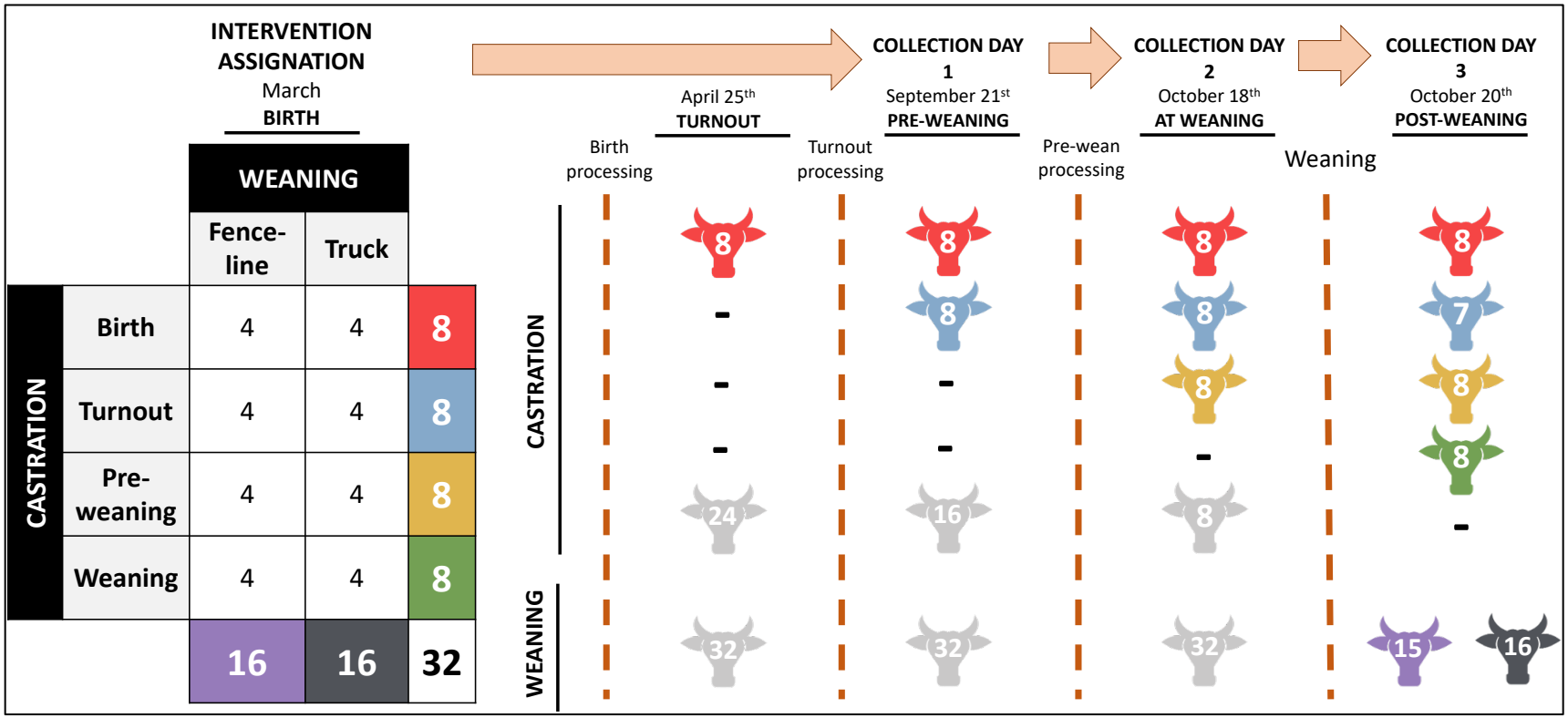
* Human (*homo sapiens*) reads accounted for 0.022% of classified reads.

Supplementary Table 2.4. List of significantly differentially abundant genera between fence-line and truck weaned calves at post-weaning day (reference group: fence-line). Only genera with logFC > 0.5 or < -0.5 and average abundance > 5 are listed.

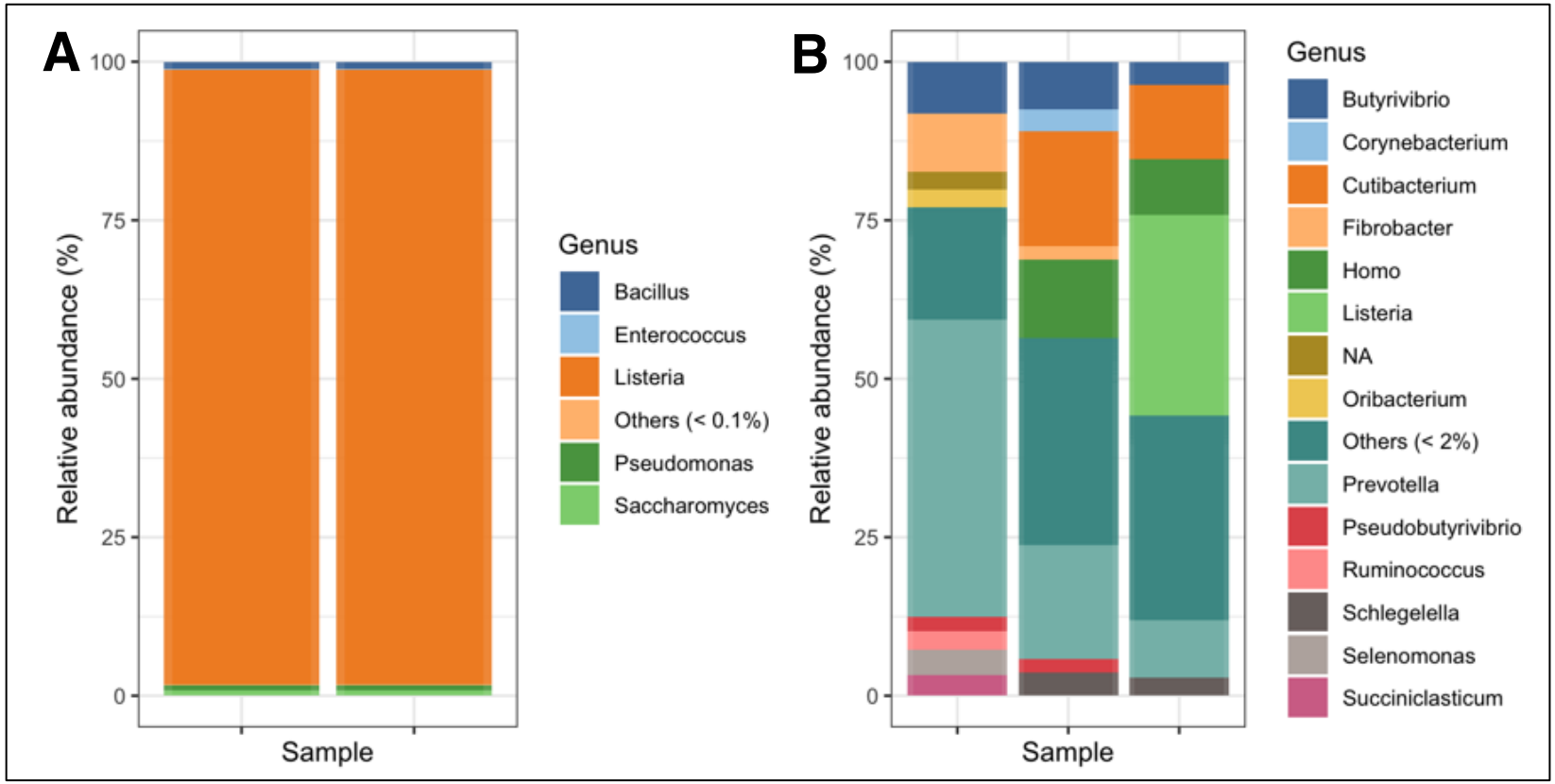
Genus	logFC	Average abundance	adj.P.Val
Rhodococcus	2.09	6.86	1.17E-15
Agrobacterium	3.14	7.16	4.46E-13
Anaerovibrio	2.77	7.90	1.52E-09
Oribacterium	2.43	12.15	7.35E-13
Plantibacter	2.42	5.87	9.91E-13
Lachnoclostridium	2.14	9.30	3.09E-10
Variovorax	2.52	5.21	1.32E-11
Mogibacterium	1.26	5.56	9.50E-11
Selenomonas	1.99	12.67	1.69E-09
Microbacterium	1.86	7.98	3.92E-12
Eubacterium	1.60	10.15	1.53E-08
Arthrobacter	0.91	5.49	9.17E-09
Succiniclasticum	1.28	13.40	1.49E-04
Stenotrophomonas	0.99	7.24	6.40E-04
Mycolicibacterium	1.06	6.50	2.73E-07
Butyrivibrio	0.95	14.36	1.30E-06
Berryella	0.76	5.22	6.61E-06
Nocardioides	1.03	6.62	8.16E-06
Pseudomonas	0.72	9.66	1.76E-02
Rhizobium	0.80	5.55	2.12E-05
Sodaliphilus	0.75	6.02	3.47E-05
Methanobrevibacter	0.71	11.41	4.71E-04
Bradyrhizobium	0.96	6.23	8.03E-05
Vescimonas	0.55	7.31	5.93E-04
Lachnobacterium	0.73	6.20	5.24E-03
Kandleria	1.65	6.96	6.15E-03
Methylorubrum	0.56	5.27	1.06E-02
Methylobacterium	-0.69	8.50	3.88E-03
Bacteroides	-1.01	11.27	6.18E-05
Lachnospira	-1.06	7.76	7.29E-03
Petrimonas	-0.56	5.00	8.60E-03
Micromonospora	-0.53	6.09	1.05E-02
Fibrobacter	-1.12	14.26	1.23E-02
Sarcina	-1.21	12.07	9.07E-06
Ruminococcus	-1.38	13.07	3.80E-08
Streptococcus	-2.31	9.85	6.41E-13

Supplementary Table 2.5. List of significantly differentially abundant methanogenic genera and genes between fence-line and truck weaned calves at post-weaning day (reference group: fence-line). Only features with logFC > 0.5 or < -0.5 and average abundance > 5 are listed.

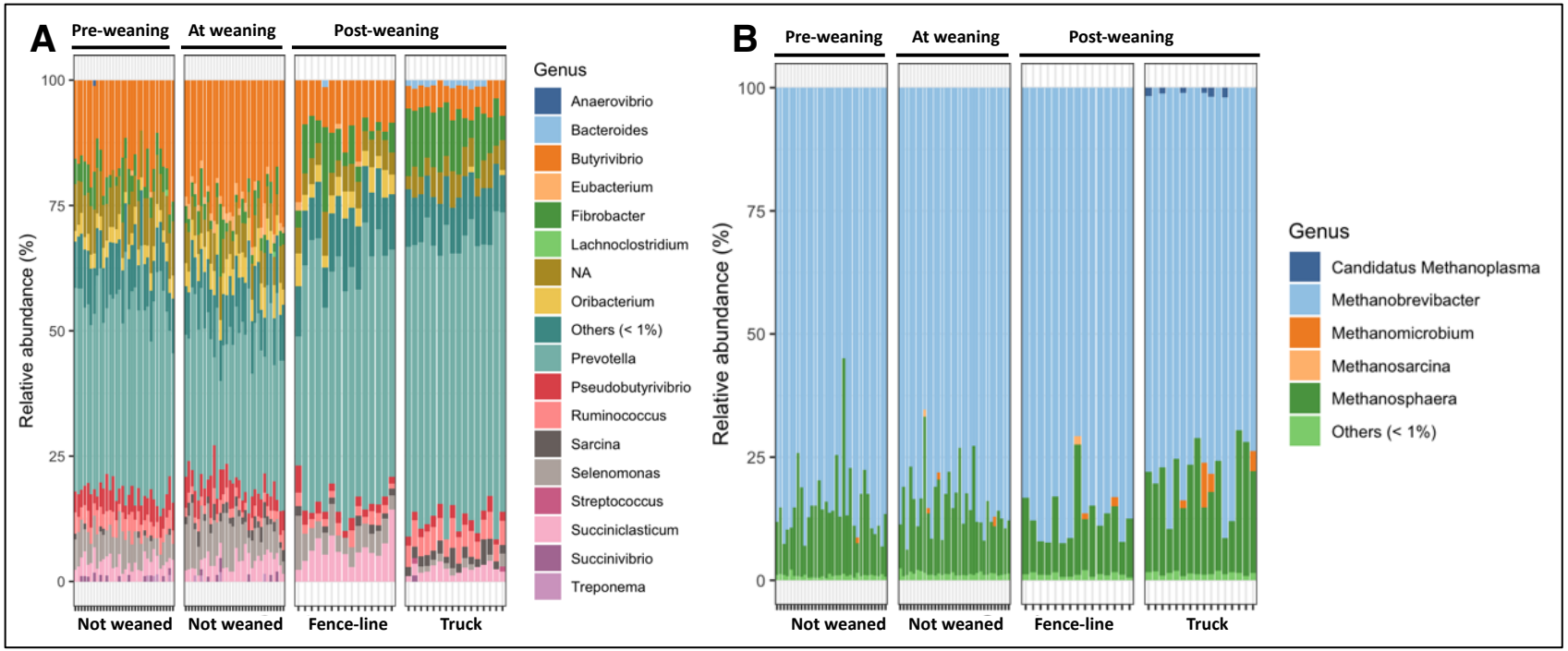
Feature		logFC	Average abundance	adj.P.Val
Genera				
	Methanobrevibacter	0.78	10.68	1.2E-05
	Selenomonas	0.82	7.74	3.4E-06
	Fibrobacter	-1.19	8.50	6.2E-04
	Candidatus Methanomethylophilus	-0.82	5.17	2.6E-03
	Ruminococcus	-0.59	7.58	3.3E-05
Genes				
	echB	0.49	5.62	1.6E-04
	fwdB	0.46	5.51	2.6E-04
	echA	0.41	6.37	5.4E-04
	fwdF	0.49	5.67	1.5E-03
	mtmB	-1.40	5.05	1.0E-07
	mtaB	-0.87	5.29	4.4E-08
	glpX	-0.60	7.51	5.5E-06
	pqqF	-0.55	8.02	3.5E-05
	fdhA-K00148	-0.55	7.94	2.0E-05
	fpoJ	-0.53	5.50	4.7E-04
	metF	-0.51	9.45	1.1E-04
	fpoC	-0.50	5.61	6.2E-06
	rnfD	-0.50	10.38	4.7E-05



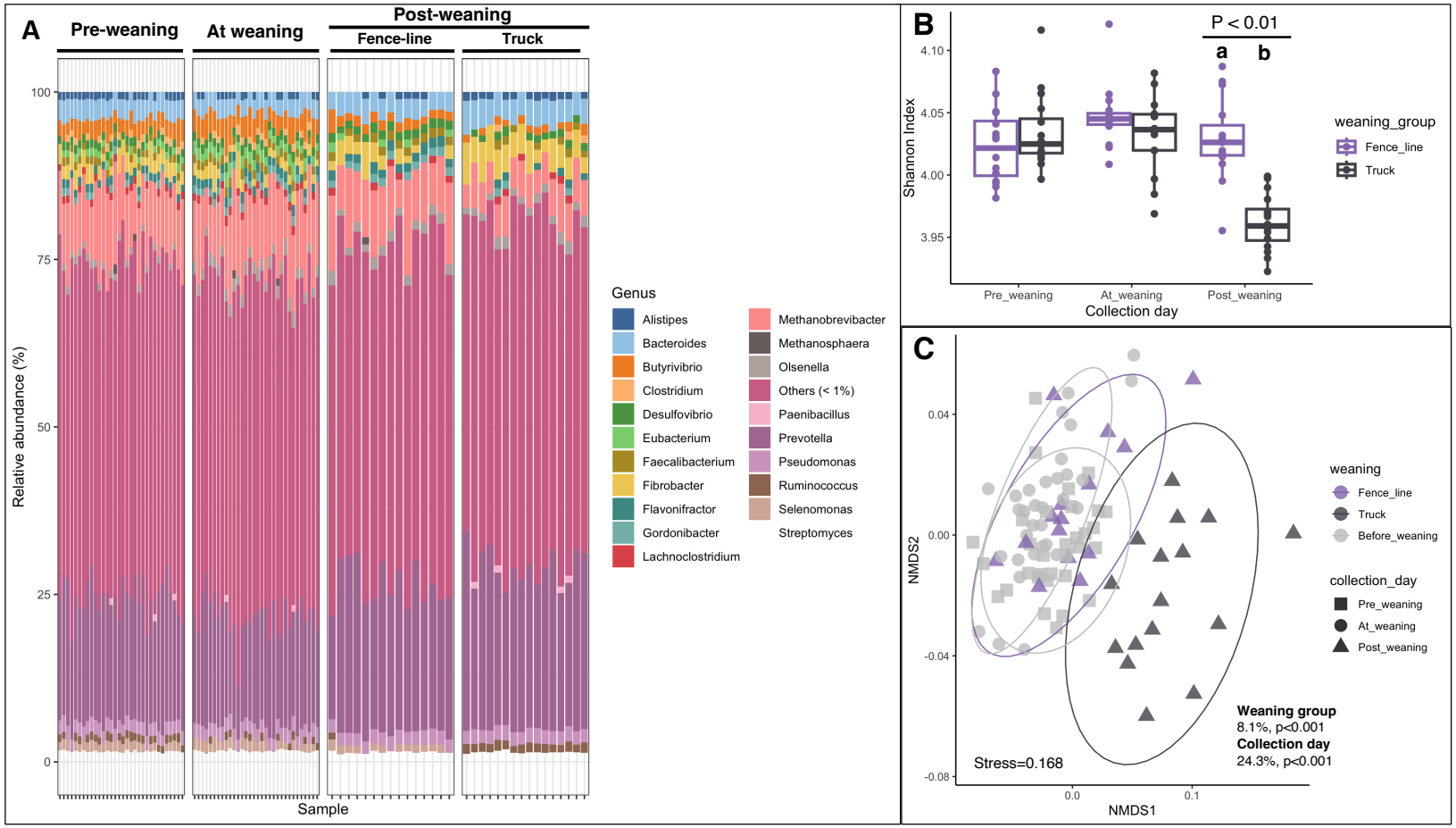
Supplementary Figure 2.1. Study design. Castration and weaning interventions per collection day, dashed lines indicate the procedure performed, numbers in white indicate the sample size for each group, each color indicates an intervention (red: birth castration, blue: turnout castration, yellow: pre-weaning castration, green: weaning castration, purple: fence-line weaning, black: truck weaning, grey: no intervention).



Supplementary Figure 2.2. Relative abundance of genus-level microbiome composition for A) Positive and B) Negative controls. Most abundant species are shown, and species < 1% abundance are grouped as “Others”. Each bar represents a single sample.



Supplementary Figure 2.3. Relative abundance plot of genus-level bacterial (A) and archaeal (B) composition, grouped by weaning strategy and collection day. Genera with < 1% abundance are grouped as "Others". Each columns represents a single sample.



Supplementary Figure 2.4. Methanogenic genera and gene content in the rumen, stratified by collection day and weaning strategy. A) Relative abundance plot of genus-level methanogenic microbe composition grouped by collection day and weaning strategy. Taxa with < 1% relative abundance are grouped as "Others". Each bar corresponds to an individual sample. **B) Alpha and C) Beta diversity plots at the methanogenic gene level.** Shannon Index is depicted as an alpha diversity index (Boxes represent the 25th to 75th percentile; horizontal line represents the median; and whiskers indicate 1.5× the interquartile range, p-values from Type-III ANOVA). Beta diversity is depicted in a Non-Metric Multidimensional Scaling (NMDS) ordination plot based on Bray–Curtis distances *(P* and R2 values from PERMANOVA testing).

CHAPTER 03: COMPARISON OF 16S rRNA GENE SEQUENCING AND DEEP SHOTGUN METAGENOMIC SEQUENCING FOR TAXONOMIC PROFILING OF THE RUMEN MICROBIOME

Gerardo R. Diaz[1], Noelle R. Noyes[1]*

[1] Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, 55108, USA

* Corresponding author: Noelle R. Noyes (nnoyes@umn.edu)

Authors' email addresses:

Gerardo R. Diaz diazo005@umn.edu

Noelle R. Noyes nnoyes@umn.edu

This chapter is in preparation to be submitted to a peer-review journal

ABSTRACT

Microbiome studies rely heavily on the sequencing approach used. Currently, 16S rRNA sequencing (16S) and shotgun metagenomic sequencing (SMS) are the most commonly used approaches. Despite well documented trade-offs between each of these sequencing approaches, they have not been formally compared for taxonomic profiling of samples obtained from the rumen of beef cattle. Rumen microbiome research is moving rapidly from 16S to SMS, particularly due to the large proportion of key non-bacterial microorganisms that cohabit the unique environment of the rumen and the limited resolution that 16S offers to characterize them. To better interpret the information gained from 16S and SMS rumen data, it is important to understand how they compare to one another, and specifically whether results from 16S are replicable using SMS, or vice versa. To answer this question, we used rumen fluid samples collected from beef cattle previously used in chapter 02. These samples were subjected to SMS and 16S-V4 workflows, and we compared the performance of the resulting data for taxonomic profiling, as well as for concordance of results. Although we observed discrepancies in taxonomic detection, resolution and classification, all methods led to similar ecological inferences. The inherent technical nature of each sequencing method and inconsistent databases explained the taxonomic discrepancies, while high correlation between genus abundance and moderate correlation between Shannon's indices and dissimilarity matrices may explain why ecological inferences were similar despite these technical discrepancies.

Keywords: 16S rRNA gene, shotgun metagenomics, rumen, microbiome, taxonomic profiling, comparison

BACKGROUND

The study of microbial communities, i.e., microbiomes, has increased in the past decades due to the association of microbiomes with important physiological outcomes in humans [136], animals [137] and the environment [138]. The development of metagenomics has overcome the limitations of conventional microbiology by using the genomic material of the entire microbial community to assign microbial taxonomy without need of bacterial isolation. The most prominent culture-independent approaches are 16S rRNA gene sequencing and shotgun metagenomic sequencing. The former uses targeted amplification of the 16S rRNA bacterial gene, while the latter analyzes DNA fragments from all genetic material in a given sample [139]. It is important to carefully consider the advantages and limitations of each method in the context of the specific research question being investigated in a given study.

16S rRNA gene sequencing is based on PCR amplification and sequencing of the 16S rRNA gene, which is highly conserved in bacteria and archaea. The 16S gene contains 9 hypervariable regions that can provide enough information to classify bacteria and archaea to the genus level, depending on the hypervariable region(s) used [129]. The resulting sequences are usually denoised by clustering similar sequences into amplicon sequence variants (ASVs), and removing false ASVs using a Bayesian model that accounts for PCR-induced errors [140]. This method is cost-effective because with a relatively low sequencing depth we can capture most of the taxonomic diversity in a given sample, at least at the genus or higher classification levels [141]. In addition, the use of ASVs does not necessitate a reference database, thus enabling the identification of uncharacterized microorganisms [142]. However, the method's effectiveness can vary depending on which hypervariable region of the 16S rRNA gene is targeted, as some regions are more variable than others [143]. In particular, the use of only one hypervariable region typically limits the ability to distinguish between different species. Additionally, the 16S method can only provide taxonomic information for bacterial and archaeal organisms (and not other microorganisms such as viruses, fungi or microscopic eukaryotes) [144]. The ability to infer functional potential from 16S bacterial taxonomy is also severely limited [145]. Dynamic evolutionary mechanisms such as horizontal gene transfer, gene duplication or gene loss can hinder an accurate functional prediction based on taxonomy because function is not always associated to a given taxon [146,147]. Functionally divergent

species within a genus, such as *Prevotella* [92], can be overlooked when only genus-level taxonomy is considered to predict functionality.

Shotgun metagenomic sequencing involves random fragmentation of all the DNA contained within a sample, coupled with short-read sequencing of the fragments. Analysis of shotgun metagenomic sequence data involves computationally-intensive tools that can assign microbial taxonomy to sequencing reads or to the longer fragments, i.e., contigs, obtained by assembling these sequencing reads [139]. Most shotgun metagenomic taxonomic profiling tools use a reference database to classify the shotgun sequencing reads based on one of 3 approaches: DNA-to-DNA (e.g., Kraken 2), DNA-to-protein (e.g., Kaiju), and DNA-to-marker-gene (e.g., Metaphlan2) [148]. In selecting an approach, one should consider the impact of the database used for the classification and the tradeoffs on precision-recall and computational requirements [149]. For instance, Metaphlan 2 [150] has high precision with shallow sequencing data and low computational requirements but lacks a customizable database; Kaiju [151] has the highest precision but also the highest computational requirements; while Kraken 2 [114] has the highest recall and customizable database but low precision [148,149]. Due to its untargeted nature, shotgun metagenomic data can be used to classify organisms other than bacteria and archaea with a resolution up to the species or even strain level [152]. The generated DNA sequence dataset can also be used to infer putative function and reconstruct bacterial genomes, known as metagenome-assembled-genomes (MAGs) [48]. However, these analyses require a higher sequencing depth (and thus cost) to accurately represent all the organisms in a microbial community. Moreover, the accuracy of taxonomic and functional profiling depends on the reference database used, which can lead to biased classification towards genomes and functions that are overrepresented in databases [148].

Previous microbiome studies have compared 16S rRNA gene sequencing and shotgun metagenomic approaches in human stool [141,153], soil [154] and the chicken gastrointestinal tract [155]. The results from these studies showed that shotgun metagenomic sequencing can be used to detect most of the taxa identified by 16S sequencing, as well as additional, biologically meaningful low-abundant taxa not identified by 16S sequencing with a higher resolution power. This observation suggests that sequencing depth and the ability of the method to resolve taxa are critical factors to fully represent the diversity of a microbial community.

These considerations are especially germane to the rumen microbiome, which is known to be a complex and rich microbial community with a key role in feed conversion and methane production in cattle [7,156]. However, no study has systematically compared the capability of these two methods for taxonomic profiling of the rumen microbiome. This is a problem because many of the earlier rumen microbiome studies based their conclusions on the 16S rRNA gene sequencing approach [9,15,25,157], whereas in recent years, shotgun metagenomics has become an increasingly popular approach to study the rumen microbiome [33,61,158]. The lack of clarity regarding the similarities and differences between methods impedes the ability to interpret and compare results between existing rumen microbiome studies. This ambiguity can confound discernment of whether differences in results arise from biological variation or discrepancies in methodologies, leading to inconsistencies in the scientific literature. A deeper evaluation of both approaches is required to understand their consistencies and inconsistencies, and to leverage the synergism of both methods to study the rumen microbiome.

Our objective was to help fill this scientific gap by undertaking a formal comparison of 16S and shotgun metagenomic rumen datasets. We hypothesized that taxonomic classification of paired rumen samples using 16S rRNA and shotgun metagenomic sequencing would lead to similar conclusions about microbiome diversity in the rumen. To test our hypothesis, we used ruminal fluid DNA samples collected from our previous study that described the effect of weaning strategy on the rumen microbiome of beef calves. These samples were subjected to both shotgun metagenomics and 16S rRNA gene sequencing and the resulting datasets (i.e., 16S and shotgun metagenomic) were compared in terms of (1) performance for taxonomic profiling of the rumen microbiome, and (2) concordance in terms of taxonomic classification and conclusion inferences about microbiome dynamics.

RESULTS

Discrepancies in taxonomic resolution and taxa identification

We compared results of taxonomic classification of beef calf rumen samples using 16S rRNA gene hypervariable region V4 sequencing ("16S-V4") and shotgun metagenomic sequencing ("SMS"). A total of 93 rumen fluid samples from 32 beef calves collected at pre-weaning, weaning and post-weaning (see chapter 02) were used to compare both sequencing approaches. After trimming and pre-processing steps, the 16S-V4 reads were classified using *dada2* v1.22 [140] with default parameters, while the SMS reads were classified using *kraken2* [114] with the minimum confidence score (cs = 0, "SMS cs0") and the maximum confidence score (cs = 1, "SMS cs1"). On average, 16S-V4 produced 96 x 10^3 paired-end (PE) reads per sample; after the trimming process, 61.4 x 10^3 PE sequencing reads per sample (63.9% of raw reads) remained for taxonomic classification. SMS produced on average 54.2 x 10^6 PE reads per sample, but only 5.8 x 10^6 PE sequencing reads per sample (10.8% of raw reads) were classified with SMS cs0, and 52 x 10^3 PE sequencing reads (0.1% of raw reads) with SMS cs1 **(Table 3.1)**. From the total number of classified reads, 98.7%, 89.3% and 77.6% were resolved to the order level by 16S-V4, SMS cs0 and SMS cs1, respectively. At the genus level, this pattern was inverted, with the proportion of classified reads lowest for 16S-V4 (70%) compared to SMS cs0 (83.9%) and SMS cs1 (72%). Likewise, the number of unique taxa identified across all samples at the order level was similar for 16S-V4 (n=178), SMS cs0 (n=173) and SMS cs1 (n=156) but was lower for 16S-V4 at the genus level (n=403) compared to SMS cs0 (n=1145) and SMS cs1 (n=861) **(Figure 3.1).**

Across all samples, the 16S-V4 method identified a relative abundance > 1% for the Patescibacteria, Spirochaetota and Verrumicrobiota phyla, while SMS cs0 and SMS cs1 identified these phyla as low abundance (i.e., < 1%). Another important discrepancy in relative abundance at the phylum level was with Proteobacteria, Actinobacteria and Firmicutes; specifically, 16S-V4 identified Proteobacteria and Actinobacteria in lower relative abundance and Firmicutes in higher relative abundance compared to both SMS methods. Noticeably, SMS cs0 identified Proteobacteria and Actinobacteria phyla in higher relative abundance compared to SMS cs1 **(Supplementary figure 3.1)**. At the genus level across all samples, 16S-V4 estimated the relative abundance of *Olsenella, Lachnospiraceae, Succiniclasticum, Rikenellaceae, NK4A214 group* (family Ruminococcaceae) and *Ruminococcus* genera at > 3% each, while the

SMS methods identified the relative abundance of these genera < 3%. SMS cs1 identified across all samples the *Fibrobacter, Butyrivibrio* and *Prevotella* genera in higher relative abundance compared to the other 2 methods, while the *Methanobrevibacter* genus was identified in higher relative abundance across all samples by 16S-V4 compared to the SMS methods. Strikingly, SMS cs0 identified a much higher proportion of low-abundant genera (defined as genera with <3% relative abundance) across all samples compared to the other 2 methods **(Figure 3.2)**.

We investigated the number of taxa commonly identified by the 3 methods and by each method individually, at the phylum and genus levels. At the phylum level, we identified 13 phyla with discordant names in the reference databases for 16S-V4 versus SMS, even though these differently-named phyla were actually the same taxon **(Supplementary table 3.1)**. After correcting this naming disagreement, we identified 18 phyla that were detected by all 3 methods, 11 phyla detected solely by 16S-V4 and 6 solely by SMS cs0 **(Supplementary figure 3.2)**. At the genus level, we noticed that there were discrepancies not only with phylum-level names but also with order-level names for some genera. Even when the genera had the same genus-level name, they were classified differently at phylum and order levels **(Supplementary table 3.2)**. Thus, we kept only the genus name (ignoring other taxonomic information), and then agglomerated the full matrix to obtain the genera that were identified by multiple methods (i.e., 16S-V4, SMS cs0 or SMS cs1), and the genera that were identified by only one of the methods. From 403 genera identified by 16S-V4, 148 were also identified by SMS cs0 and SMS cs1, while 253 were exclusively identified by 16S-V4. From 1145 genera identified by SMS cs0, 2 were also identified by 16S-V4, 713 were also identified by SMS cs1, while 282 were exclusively identified by SMS cs0 **(Figure 3.3).** The 148 genera identified by all 3 methods represented 45% (148/403) of total genera identified by 16S-V4 across all samples, 13% (148/1145) of genera identified by SMS cs0, and 17.2% (148/861) of genera identified by SMS cs1. However, the proportion of total classified reads represented by these 148 genera had the opposite pattern across the 3 methods; namely, they represented 29.3% ($1.7/5.8 \times 10^6$), 49.5% ($266.7/538.5 \times 10^6$) and 64.6% ($3.1/4.8 \times 10^6$) of total classified reads across all samples for 16S-V4, SMS cs0 and SMS cs1 method, respectively.

Using the 148 genera identified by all 3 methods, we analyzed the pairwise Pearson's correlation of genera abundance across the 3 methods and found them to be significantly

different **(ANOVA $P < 0.05$, Figure 3.4).** The genus-level abundances in the SMS cs0 and SMS cs1 datasets had the highest average correlation values (R^2= 0.94, min= 0.79, max=0.99), while the abundances from 16S-V4 and SMS cs1 had the lowest (R^2= 0.85, min=0.59, max=0.97). The average correlation value for 16S-V4 and SMS cs0 abundances was 0.89 (min= 0.58, max=0.96).

Different but correlated diversity estimates led to similar inferences about rumen dynamics

Using shotgun metagenomic data and a kraken confidence score of 0.1, we previously concluded that both the act of weaning and weaning strategy are significantly associated with changes in rumen alpha and beta diversity (see chapter 02). In the current analysis, we evaluated whether diversity estimates obtained from the 16S-V4, SMS cs0 and SMS cs1 methods would lead to similar inferences regarding rumen dynamics at weaning. We obtained the alpha and beta diversity metrics separately for each method. Overall, SMS cs0 had the highest genus-level Shannon's index values across all samples (mean= 5.04, min=4.55, max=5.34), while cs0 had the lowest (mean=1.93, min=1.29, max=2.66). The association of the genus-level Shannon's index with both weaning strategy and collection day was statistically significant using results from all 3 methods. Despite differences in the point estimates, the Shannon's index was significantly different between post-weaning samples collected from calves weaned by fence line (adjusted means ± SE: 16S-V4= 3.92 ± 0.0336; SMS cs0= 4.93 ± 0.0354; SMS cs1= 1.81 ± 0.0590) and calves weaned by truck (adjusted means: 16S-V4= 3.8 ± 0.0325, SMS cs0= 4.83 ± 0.0342, SMS cs1= 1.58 ± 0.0571). These results are consistent with the results previously described (see chapter 02). The difference in Shannon's index (effect size) between the truck and fence-line weaning groups in the post-weaning samples was largest in the SMS cs1 results compared to SMS cs0 and 16S-V4 (0.23 versus 0.1 and 0.12, respectively) **(Figure 3.5. A-C)**.

Similarly, all 3 methods indicated a significant difference in beta diversity by both collection day and weaning strategy, which was also consistent with patterns reported previously (see chapter 02). However, estimates of the amount of variability partitioned to each variable (i.e., the R^2 values from PERMANOVA testing) were different for each method. When the genus-level 16S-V4 dissimilarity matrix was used, 20.5% and 3.1% of the between-sample variability was partitioned to collection day and weaning strategy, respectively (both PERMANOVA $P < 0.001$).

When genus-level SMS cs0 and cs1 dissimilarity matrices were used, the amount of variability partitioned to collection day increased relative to the 16S-V4 estimate, while the amount partitioned to weaning strategy remained consistent; specifically, the SMS cs0 R^2 values for collection day and weaning strategy were 30.1% and 3.1%, respectively, while for SMS cs1 these values were 31.1% and 3.3% **(all PERMANOVA *P < 0.001,* Figure 3.5. D-F).**

To explore how the diversity metrics estimated from the 3 methods led to similar conclusions despite being different in magnitude (i.e., effect size), we evaluated the correlation between alpha diversity indices and beta diversity results. The R^2 Pearson's correlation value for richness between the three methods was neither high nor statically significant, however the Shannon's index values between SMS cs0 and SMS cs1 were highly correlated (R^2= 0.88, p-value < 0.001), while the Shannon's index values of 16S-V4 were slightly more correlated with values of SMS cs0 (R^2= 0.52, p-value < 0.001) than with SMS cs1 (R^2= 0.50, p-value < 0.001) **(Table 3.2)**. When the correlation of the dissimilarity matrices obtained by each method were evaluated using Procrustes analysis, we found a statistically significant correlation between all pairwise comparisons, with correlation between SMS cs0 and SMS cs1 achieving the highest value (Procrustes m^2=0.17; correlation=0.9). The 16S-V4 dissimilarity matrix was more correlated with SMS cs0 (Procrustes m^2=0.23; correlation=0.88) than with SMS cs1 (Procrustes m^2=0.33; correlation=0.82) **(Figure 3.6)**. These results were consistent with the correlation of genera abundance and genus-level Shannon's index: 16S-V4 indices had a higher correlation with SMS cs0 than with SMS cs1, and the correlation between SMS cs0 and SMS cs1 was the highest.

DISCUSSION

As the cost of DNA sequencing steadily decreases, shotgun metagenomic sequencing (SMS) will likely become a major component of a multi-omics framework to study microbial communities. Comparison of the taxonomic profiling capabilities of 16S rRNA gene sequencing (16S) versus SMS will provide valuable insights to leverage the large extant 16S-based microbiome literature. This is relevant for rumen microbiome research because the use of SMS to understand the functional role of rumen microbes in cattle production is rapidly increasing. Using rumen fluid samples collected from a cohort of 32 beef calves at three collection days (see chapter 02) and simultaneously sequenced by SMS and 16S-V4, we compared: (1) the performance of 16S-V4, SMS cs0 and SMS cs1 for taxonomic profiling of the rumen microbiome of beef cattle and (2) the agreement between methods in terms of taxonomic classification and inferences about microbiome dynamics. We identified discrepancies in taxonomic classification resolution, number of unique taxa identified and taxonomic labels at different levels; the magnitude of the discrepancy depended on taxonomic rank, with much larger discrepancies at the genus level. Technical differences inherent to the sequencing methods and disagreement in the taxonomic information embedded in the databases used for 16S-V4 and SMS can be pointed to as possible explanations for these discrepancies. Importantly, despite these discrepancies, similar conclusions about the overall study results were inferred with the 3 methods, including consistency in statistical significance and the direction of associations. This overall concordance in results could be explained by the high correlation of abundances (i.e., counts) of commonly identified genera across all methods; as well as moderate correlation between Shannon's indices (alpha diversity); and dissimilarity matrices (beta diversity).

The agreement between 16S and SMS to infer the same conclusion from an experimental design has been widely reported in microbiome research. Regardless of the bioinformatic analysis method, sequencing depth or type of sample, results obtained from 16S and SMS datasets were generally concordant when compared in soil [154], human gut [141,153,159], chicken gut [155], horse gut [160] and even pollen [161] microbiome studies. However, a previous rumen microbiome study [49] reported inconsistent results in differential abundance testing, especially for Firmicutes and Bacteroidetes phyla when rumen samples were analyzed in parallel by 16S and SMS methods. The reasons for these contrasting results are likely related to differences in sample size, 16S hypervariable region, OTU picking method and SMS analysis

methods. Ultimately, the level of concern about method concordance depends on the research question and the biological signal that is being investigated. For instance, in our study, *Methanobrevibacter* and *Ruminococcus* genera were identified in higher relative abundance across samples by 16S-V4, while *Fibrobacter, Butyrivibrio* and *Prevotella* genera were identified in higher abundance by SMS methods. Having an specific genus-driven research question would leverage the advantage of each method to detect a given genera as each taxa is associated with different roles in the rumen microbiome [9,32]. Nonetheless, a higher signal obtained for a given genera does not always guarantee an accurate measurement of the microbial community.

The discrepancies in genus identification between 16S and SMS reported in this study are likely attributable to differences in the reference databases and intrinsic technical characteristics of each method. For example, SMS' ability to identify more low-abundant taxa and its superior classification resolution compared to 16S have been previously discussed in other comparison studies in soil [154] and chicken gut microbiome [155], suggesting that sequencing depth and the untargeted nature of SMS are important for identifying more genera and thus improving genus-level analyses. Conversely, 2 comparison studies in water [162] and pediatric fecal microbiome samples [141] identified more genera using 16S than SMS. The low sequencing depth of SMS (i.e., 10-12M reads per sample) and the use of a marker-gene-based SMS classifier with non-customizable database (i.e., MetaPhlan2) in those two studies may explain the contrasting findings.

An important constraint that we experienced in the current study is the inconsistent taxonomic labels in the taxonomic classification schemes used by 16S (e.g., SILVA) and SMS (e.g., RefSeq NCBI). These inconsistent taxonomic labels prevented a more comprehensive comparison between the methods. The need for a unified taxonomy for all databases has also been pointed out by other comparison studies [141,153,160]. Some solutions have been proposed, such as the use of cross classification across databases using a bioinformatic tool [163], and the use of a candidate 16S-SMS-harmonized database [153]. Regarding the taxonomic profiling of rumen microbiome by SMS, we already showed the impact of the reference database in taxonomic classification rate (see chapter 02) consistent with other studies [46,48].

The prevalence and abundance of the 148 genera identified by all three methods (i.e., 16S-V4, SMS cs0 and SMS cs1) exhibited contrasting patterns across the three approaches. While these genera represented a high proportion of total genera (45%), but low proportion of total reads classified by 16S-V4 (29.3%), the opposite was true for SMS methods, i.e., these same genera represented a low proportion of total genera (13% and 17.2%) but a high proportion of total classified reads (49.5% and 64.6%). We did not expand on the relevance of genera commonly and exclusively identified by different methods in our study, but a recent comparison study showed that unlike 16S-exclusive genera, the genera solely identified by SMS could reflect key diversity differences from the experimental design [155]. This result and the differences in genus-level relative abundances estimated by 16S and SMS, brings to discussion the potential pitfalls of the 16S method that could explain our findings. Several sources of bias have been shown to impact the performance of 16S rRNA sequencing, such as different 16S rRNA gene copy number in bacteria [164], the polymerase, primers and amplification cycles used in the 16S PCR [165,166], and the methods used for clustering or denoising 16S sequences during bioinformatic analysis [167]. It is important to mention that our study was limited to the commonly used V4 hypervariable region of 16S rRNA gene, but the use of full-length 16S gene sequencing (~1.5 Kbp) may provide improved taxonomic resolution for many sequences, thus ameliorating the discrepant diversity results corresponding to different 16S hypervariable regions [168,169].

Despite the taxonomic and functional profiling advantages of SMS [139], this method also contains sources of bias, including those related to the library preparation process (e.g., DNA fragmentation, library amplification, polymerase performance and GC% content of sequenced genomes) [170]; the proportion of host DNA in a sample [171]; reference databases; and the technical tradeoffs of choosing different SMS taxonomic classifiers [148,149]. The advantages of microbiome research are widely known but their limitations are often poorly addressed. Regardless of the sequencing method, microbiome studies present several sources of bias which are beyond the scope of this study but are well discussed elsewhere [41,172,173]. This study was not designed to compare our results with a ground truth, thus we cannot rule out other potential sources of bias that can be adding variability to our 16S-SMS comparison.

Remarkably, despite all these methodological constraints and the numerous discrepancies in taxonomic results, all 3 methods that we compared led to similar overall conclusions about the

dynamics of the rumen microbiome in beef calves. We provided evidence that this congruence in results may be due to relatively robust correlation of diversity metrics across the three methods. Our observation that Shannon's index obtained by SMS cs0 was higher than 16S-V4 has been previously reported, suggesting that the much higher sequencing depth of SMS may help to detect genera that are either very low-abundance and/or not resolvable at the genus level by 16S rRNA sequencing [154,159,160]. Likewise, the moderate correlation we observed between Shannon's indices calculated by 16S and SMS methods was consistent with previous studies. A moderate-to-high correlation (R^2= 0.63) between Shannon's indices obtained by 16S-V4 and SMS was reported in human gut samples [159], while a moderate (R^2= 0.46) but still statistically significant correlation in Shannon's indices obtained by 16S-V4 and shallow SMS was reported in samples collected from the equine gut [160]. The significant correlation between Bray-Curtis dissimilarity matrices calculated with 16S and SMS was also supported by a recent study of the horse gut microbiome [160].

Kraken 2 is a widely used method for taxonomic profiling of metagenomic short-read data. Both versions, kraken [174] and kraken2 [114] account for a total of 5801 citations registered in PubMed (accessed on May 2023). Because Kraken 2 favors recall over precision [149], the developers provide a way to control for potential false positive results, i.e., the confidence score. This parameter sets the stringency level used in the k-mer-based classification using a lowest-common ancestor (LCA) approach [114]. However, little information or recommendation is provided by the authors on how to use this confidence score and the impact that it may have on microbiome studies. In the same way, studies that use Kraken 2 do not typically provide details on the confidence score used, or the reasons for its selection. For instance, from 22 rumen microbiome studies that cited kraken or kraken 2 as taxonomic classifiers according to PubMed (accessed on May 2023), only 2 detailed the confidence score that they used for the taxonomic classification [40,175]. We compared the 2 most extreme confidence scores and noticed a clear impact on classification rate, which was further reflected in the number of genera detected as well as the relative abundance of those genera across samples and the diversity metrics obtained from those genera. Comparing results obtained from a confidence score of 0 versus 1, we observed an ~100-fold decrease in classified reads **(Table 3.1)**; 279 fewer genera; and 1,598 fewer species across all samples **(Figure 3.1)**. These decreased classifications led to Shannon's indices decreasing by approximately half **(Figure 3.5).** However, the cs0 and cs1 methods showed the highest correlation in genus-level abundances, as well as alpha **(Table**

3.2) and beta diversity metrics **(Figure 3.6)**. A recent study comprehensively compared all possible kraken 2 confidence scores with different databases and sample matrices and provided general recommendations for setting an appropriated confidence score, which we adhered to in our previous study of rumen microbiome dynamics (see chapter 02). After performing an initial exploratory analysis, the authors suggested to decrease the confidence score under the following scenarios: for samples collected from a poorly explored source; when a high level of microbiome diversity is expected; or when false negatives are a concern. Conversely, the authors suggested increasing the confidence score if samples are from a well-studied environment; when diversity is expected to be low; and if false positives are a concern [176].

Finally, based on the results of this study we make several suggestions for future rumen microbiome studies. First, the decision of which sequencing approach to use should be made based on the research question. Depending on the hypothesis, researchers may be more interested in performing taxonomic profiling of bacteria or archaea at less-resolved levels of the taxonomy (i.e., order, class or phylum), in which case 16S-V4 may be preferable. However, if taxonomic domains other than bacteria and archaea are desired; or if functional profiling or strain-level information is needed; then SMS is a better option. Second, logistical constraints such as available budget, sequencing and computational resources will likely influence the selection of methodology. Generally speaking, SMS is more costly and requires larger sequencing platforms and computational resources than 16S sequencing. Third, in order to make future rumen microbiome results more readily comparable and reproducible, authors are encouraged to provide as much detail as possible on the methodology, for instance: (a) the sequencing depth, (b) the database used for taxonomic classification, (c) detailed information about the setup and parameters used with the bioinformatic tools, (d) how any sequencing reads were dropped and retained at each step in the bioinformatic workflow, (e) the number of unique taxa and percentage of sequencing reads classified at each given taxonomic rank, and (f) details on statistical methods used for the analysis. Ideally, these details would also be captured in reproducible code files, with the corresponding raw data in a repository.

CONCLUSION

This study compared 3 approaches for taxonomic classification of the rumen microbiome of beef calves: 16S rRNA gene hypervariable region V4 sequencing, shotgun metagenomic sequencing using *Kraken2* with confidence score 0 and with confidence score 1. While the methods produced discordant results in terms of numbers of unique taxa detected, taxonomic resolution and diversity estimates, the overall conclusions inferred from the experimental design of the study were similar. Overall, this study highlights the importance of carefully selecting the taxonomic classification approach for rumen microbiome studies, while balancing the research question(s) with budgetary and resource constraints. Finally, we emphasize the need to publish a reproducible and detailed methods protocol, which will support future comparative studies, including systematic reviews and meta-analyses.

METHODS

Overview

Using rumen fluid samples from our previous study (see chapter 02), we compared the performance of 16S rRNA gene sequencing (hypervariable region V4) and shotgun metagenomic sequencing for taxonomic profiling of the rumen microbiome. The same aliquot of extracted DNA for each sample was used by apportioning ~20uL to 16S rRNA gene sequencing and ~20uL to shotgun metagenomic sequencing. Additionally, 3 negative controls (extraction blanks) and 2 positive controls were submitted for both sequencing methods. 16S rRNA gene sequences were processed and taxonomically profiled using *cutadapt* v4.0 [177], *dada2* v1.22 [140] and the SILVA nr99 v138.1 database. Shotgun metagenomic sequences were processed using the *minor_kraken2.nf* module of AMR++ v2 [74], setting 2 confidence scores for taxonomic profiling: 0 (cs0, lowest stringency) and 1 (cs1, highest stringency). *Kraken2* used the refseq database updated in July 2020 (NCBI) as the reference database. One count matrix was built with 16S-V4 output, while 2 count matrices (1 per each confidence score) were built with SMS output. We compared the 3 matrices in terms of (1) method performance for taxonomic profiling of the rumen microbiome, and (2) agreement of the methods to infer conclusions based on microbiome diversity from the experimental design of our previous study.

Sample collection and DNA extraction

In a previous study, we identified rumen microbiome changes in beef calves associated with collection day and weaning strategy, using shotgun metagenomic sequencing. The study had a randomized longitudinal factorial design, with 32 calves randomly assigned to 2 different weaning strategies and ruminal fluid collected twice before weaning and once 48 hours after weaning. A total of 95 rumen samples were collected and DNA was extracted using the Dneasy® PowerSoil® Pro Kit (QIAGEN, USA), following manufacturer instructions as previously described (see chapter 02). A single aliquot of DNA extracted from each ruminal fluid sample was submitted for 16S rRNA gene sequencing, targeting the V4 hypervariable region (16S-V4), and shotgun metagenomic sequencing (SMS) to the University of Minnesota Genomics Center (UMGC). Three negative DNA controls (extraction blanks) and 2 DNA positive controls (ZymoBIOMICS Microbial Community Standard II Log Distribution – Catalog N° 6310) were submitted alongside the samples.

16S-V4 sequencing and analysis workflow

DNA samples were submitted for qPCR of the 16S rRNA gene and subsequent normalization of the estimated initial bacterial DNA concentration for library preparation. The 16S rRNA library preparation targeted the V4 hypervariable region of bacterial 16S rRNA gene using a dual-indexing 16S rRNA Illumina primer set: Forward 5' -GTGCCAGCMGCCGCGGTAA- 3'; Reverse 5' -GGACTACHVGGGTWTCTAAT- 3' [178]. Paired-end sequencing was performed on the MiSeq platform (Illumina, USA) using Illumina v3 cluster chemistry with PhiX control v3 to produce 2x300 bp paired-end reads.

Demultiplexed sequences were processed and analyzed using appropriate packages in R Statistical Software (v4.2.1; R Core Team 2022; https://www.r-project.org/). V4-region primers were removed from the sequencing reads using *cutadapt* v4.0 [177]. Sequence trimming, denoising and taxonomy assignment were done using *dada2* v1.22 [140]. Sequence trimming was based on the quality profile plots of sequencing reads, considering a truncated quality score ≤ 2, truncated sequence length of 240 bp for forward and 160 bp for reverse reads, and a maximum expected error (maxEE) of 4,4. PhiX control sequencing reads and sequencing reads with ambiguous bases (N) were removed. The denoising process of trimmed reads was performed with dereplicated sequences, using an error rate estimated with a Bayesian model that accounts for PCR-induced errors in our specific dataset [140]. The denoised paired-end sequences were merged and the sequence artifacts (sequences with length less than 251bp and more than 255bp) and chimeras were removed to obtain the Amplicon Sequence Variants (ASV). Taxonomic assignment of ASVs was done by the naïve Bayesian classifier method [179] using the SILVA reference database nr99_v138.1 up to the genus level, and the SILVA species-level database species_assignment_v138.1 up to the species level. Finally, we generated an ASV count table and an ASV taxonomic table, which were combined with metadata to obtain a count matrix using *Phyloseq* v1.4 [126]. The frequency-based method was used to remove contaminant ASVs from our count matrix using *decontam* v1.18 [180]. The decontaminated count matrix was used for downstream analysis.

Shotgun metagenomics sequencing and analysis workflow

Library preparation and shotgun metagenomic sequencing were previously described (see chapter 02). Demultiplexed paired-end reads (2x150bp) were analyzed using the AMR ++ v2 pipeline [74], which integrates sequence trimming and filtering by *Trimmommatic* v0.4 [119],

host read decontamination by BWA v0.7.17 [120] and BEDTools v2.3 [181], and k-mer based taxonomic profiling by kraken2 v2.1.2 [114]. The *Bos taurus* reference genome (Genome Bos_taurus_UMD_3.1, accession number: GCA_002263795.3) was used for decontamination of host reads. For taxonomic profiling with kraken2, we used the standard genome database (updated in July 2020, containing archaea, bacteria, virus, plasmids, UniVec_Core and human genomes), setting 2 confidence scores: 0 (cs0, minimum threshold) and 1 (cs1, maximum threshold). Kraken2 is a taxonomic classifier that assigns a taxonomy label to a sequencing read based on its k-mer profile and the lowest common ancestor (LCA) scoring approach . The confidence score sets the stringency level used in the LCA scoring approach, and thus can be used to generate classifications with varying levels of confidence [114]. Briefly, the process works thus: a single sequencing read is classified when the score ratio for a given taxonomic label bypasses the confidence score threshold. The score ratio for a taxonomic label is obtained by dividing the number of k-mers that map to a given label in the database by the total number of unambiguous k-mers in the read. If several labels exceed the threshold, then the LCA is chosen (based on the taxonomic tree structure). If the LCA label is at the root of the taxonomic tree and its score ratio does not exceed the threshold, the sequence will be labeled as "unclassified" [114]. Finally, we generated an Operational Taxonomic Unit (OTU) count matrix for the results from each confidence score using *Phyloseq* v1.4 [126]. Sequencing reads classified as viruses, viroids and eukaryotes were excluded from the 2 matrices for downstream analysis.

Comparison of method performance for taxonomic profiling of the rumen microbiome

To compare the taxonomic profiling performance of 16S-V4, SMS cs0 and SMS cs1 we reported for each method: (1) the number of sequencing reads obtained through each analysis step, (2) the proportion of reads classified and number of unique taxa detected at each taxonomic rank across all samples, (3) the relative abundance of the rumen microbial genera, (4) the number of genera identified in common and individually by the 3 methods across all samples, and (5) the pairwise correlation of genera abundance in each sample as described by the 3 methods. All the analyses were performed in R Statistical Software (v4.2.1; R Core Team 2022; https://www.r-project.org/) and plots were obtained with *ggplot2* v3.4.1 package for R [130]. Two samples (one from pre-weaning and the other from post-weaning collection) were dropped from the 3 count matrices due to technical issues during 16S-V4 sequencing; specifically, the pre-weaning sample had very low qPCR concentration and number of

sequencing reads, while the post-weaning sample contained more than twice the average number of observed genera per sample. Both samples did not present abnormal results with SMS method, but they were dropped from kraken2 cs0 and cs1 count matrices to make a balanced comparison.

To estimate the proportion of classified reads and number of unique taxa across all samples at each taxonomic rank, the 3 matrices were separately agglomerated to every taxonomic rank, from Domain/Kingdom to Species, dropping unclassified taxa (NA) using *Phyloseq.* To estimate the genus-level relative abundance, we normalized each raw matrix separately by the Cumulative Sum Scaling (CSS) method using *metagenomeSeq* package v1.4 [127]. Then, each normalized matrix was agglomerated to the genus level keeping the taxa that did not get a genus-level classification (NAs) and represented as relative abundance in a bar plot using *ggplot2*. To identify and count the number of genera detected by all 3 methods across all samples, we combined the 3 genus-level-agglomerated matrices in one matrix, while maintaining the sample names and OTU/ASV identification labels. During this process, we noticed that some genera were duplicated within the combined matrix because their label names at higher taxonomic levels did not match between the 16S-V4 and SMS reference taxonomies **(Supplementary table 1)**. To circumvent this issue, we ignored the superior taxonomic rank label names and used only the genus-level label names to agglomerate the combined matrix. Using *MicrobiotaProcess* package v1.10.2 [182] in the agglomerated full matrix, we identified the genera detected by all three methods across all samples, as well as those detected by < 3 methods in combination and separately. Finally, we plotted that information as an UpSet plot using *UpSetR* package v1.4 [183]. We repeat the same procedure with a combined phylum-level-agglomerated matrix to identify the phyla detected by the 3 methods and < 3 methods. To investigate the correlation of genera abundance detected by the 3 methods for the same sample, we subsetted only genera detected by the 3 methods. Using the subset matrix we obtained the Pearson's correlation R^2 value and p-value of the comparisons of abundance detected by Kraken cs0 vs 16S-V4, Kraken cs1 vs 16S-V4, and Kraken cs1 vs Kraken cs0 for the same sample. We plotted the R^2 values for each pairwise comparison for each sample and tested if the mean pairwise R^2 value was statistically different between each category ("Kraken cs0 vs 16S-V4", "Kraken cs1 vs 16S-V4", and "Kraken cs1 vs Kraken cs0") using ANOVA and a *post-hoc* pairwise t-test adjusting the p-values by Bonferroni method.

Comparison of diversity metrics inferred by the 3 methods and agreement on experimental design conclusions

As we already reported a significant association of alpha and beta diversity with weaning strategy and collection day (see chapter 02), we investigated this association using 16S-V4 and SMS (cs0 and cs1) methods. Unlike this work, our previous study performed SMS classification using *kraken2* with a confidence score = 0.1 and a comprehensive reference database that included archaea, bacteria, virus, plasmids, human, UniVec_core, protozoa and fungi reference genomes, as well as 410 rumen-specific bacteria genomes from the Hungate project [72]. Additionally in the previous study, our removal of "host" DNA within the SMS data included removal of reads that aligned to *B. taurus* and several plant genomes that were known to dominate the diet of the sampled calves. To compare the estimates and changes in genus-level diversity inferred from the experimental design of the study using each method, we evaluated: (1) the association of Shannon's index (alpha diversity) with the interaction of weaning strategy and collection day, (2) beta diversity by weaning strategy and collection day, (3) the correlation of beta diversity distance matrices obtained by each method, and (4) the Pearson's correlation of alpha diversity indices obtained by all 3 methods.

To investigate the association with alpha diversity, we calculated Shannon's index for each genus-level count matrix using *Phyloseq* v1.4 package and tested for statistical significance with a linear mixed-effect model using *lme4* v1.1.32 [132]. We considered Shannon's index as the dependent (outcome) variable, animal ID as a random effect, the interaction of weaning strategy with collection day as an independent variable (predictor), and castration timing as a covariate. The adjusted mean of Shannon's index for each weaning strategy was obtained by least square means using *emmeans* v1.8.5 [134]. To investigate the association with beta diversity, we calculated the Bray-Curtis dissimilarity indices of the CSS-normalized genus-level matrices and tested for statistical significance using Permutational Multivariate Analysis of Variance (PERMANOVA), both implemented in *vegan* v2.4.6 [184]. The R^2 value was used to estimate the marginal amount of variation partitioned to a given dependent variable and was reported along with the associated *P*-value to determine statistical significance. To calculate the correlation between beta diversity ordinations (Bray-Curtis distance matrix) obtained via each of the 3 methods, we performed a symmetric Procrustes analysis with *vegan* v2.4.6 [184]. Before performing the analysis, the Bray-Curtis distances were transformed to Euclidean distances

using the Cailliez adjustment [185], in order to prevent negative eigenvalues. We tested the significance of the correlation from our symmetric Procrustes analysis by a permutational test of significance (*protest*) with 999 permutations, also implemented in *vegan* v2.4.6. To evaluate the correlation of genus-level alpha diversity indices obtained by the 3 methods, we calculated the Pearson's correlation R^2 values and the associated p-values for each pairwise comparison: "16S-V4 vs SMS cs0", "16S-V4 vs SMS cs1", and "SMS cs0 vs SMS cs1".

TABLES & FIGURES

Table 3.1. Sequence read summary statistics, stratified by collection day and method (mean ± SD)

Collection day	16S rRNA gene sequencing reads (x10^3)				Shotgun metagenomic sequencing reads (x10^6)			
	qPCR (x10^6 molecules/µl)	Raw	Denoised	Classified	Raw	trimmed non-host	Classified (cs=0)	Classified (cs=1)
Pre-weaning	96.3 ± 40.7	96.8 ± 37.5	62.5 ± 29.9	62.2 ± 29.7	54.7 ± 8.9	42.2 ± 6.7	5.7 ± 1	0.049 ± 0.012^a
At weaning	108.4 ± 46.9	92.9 ± 26.8	59.5 ± 22.8	59.2 ± 22.7	53.6 ± 10.7	41.3 ± 8.3	5.8 ± 1.2	0.049 ± 0.012^a
Post-weaning	99.7 ± 27.9	98.4 ± 29.1	63 ± 24.6	62.7 ± 24.4	54.2 ± 8.9	42.4 ± 6.9	6 ± 1	0.058 ± 0.013^b

*Kruskal-Wallis test, statistical significance (P<0.05) indicated by different superscript letters

Table 3.2. Pearson's correlation values for alpha diversity metrics obtained by each method

	Richness		Pielou's evenness		Shannon's Index	
	R^2	p-value	R^2	p-value	R^2	p-value
16S-V4 vs SMS cs0	- 0.03	0.76	0.35	< 0.001	0.52	< 0.001
16S-V4 vs SMS cs1	0.11	0.28	0.38	< 0.001	0.50	< 0.001
SMS cs0 vs SMS cs1	- 0.16	0.12	0.87	< 0.001	0.88	< 0.001

16S-V4: 16S rRNA gene sequencing V4 hypervariable region; SMS cs0: Shotgun metagenomic sequencing confidence score 0; SMS cs1: Shotgun metagenomic sequencing confidence score 1.

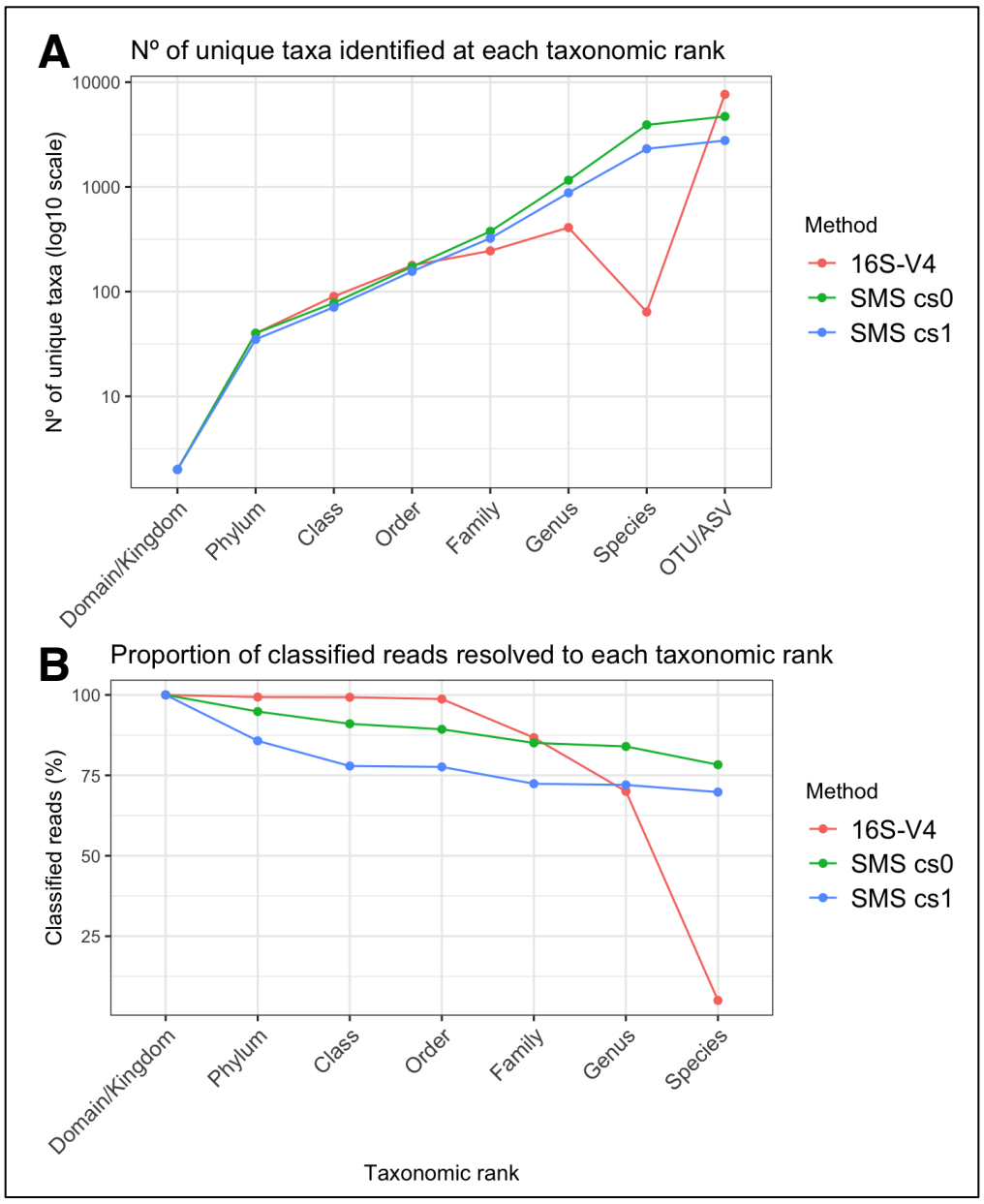


Figure 3.1. Taxonomic profiling performance of each method across all samples at different taxonomic ranks, expressed as A) number of unique identified taxa, and B) proportion of classified reads. The denominator in B represents the number of reads classified to the Domain/Kingdom level. 16S-V4: 16S rRNA gene sequencing V4 hypervariable region; SMS cs0: Shotgun metagenomic sequencing confidence score 0; SMS cs1: Shotgun metagenomic sequencing confidence score 1.

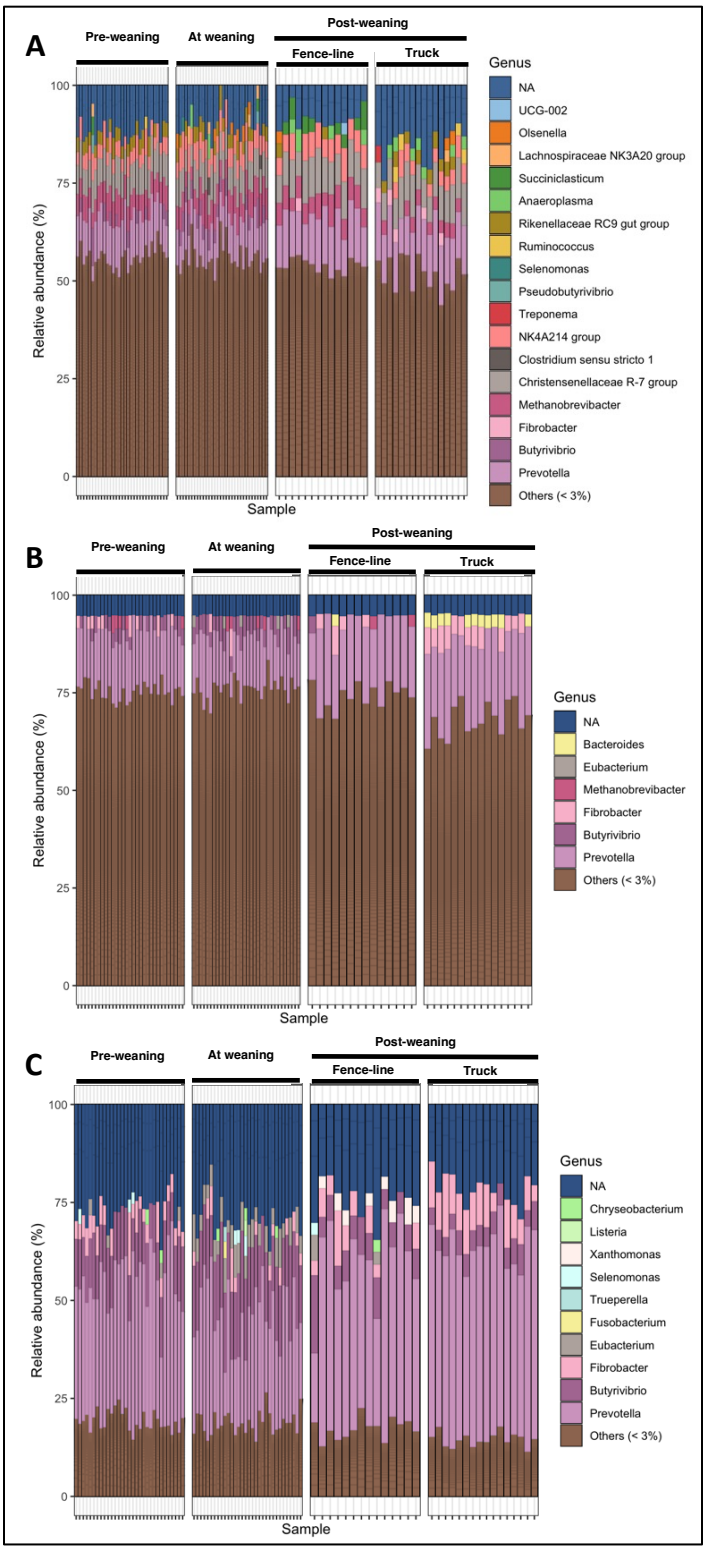


Figure 3.2 Relative abundance plots of genus-level counts obtained by A) 16S rRNA gene sequencing, B) Shotgun metagenomics cs0, C) Shotgun metagenomics cs1. Genera with < 3% relative abundance are grouped as "Others". Each bar represents one sample. NA represents taxa not classified at the genus level (i.e., they may have classification at higher taxonomic ranks)

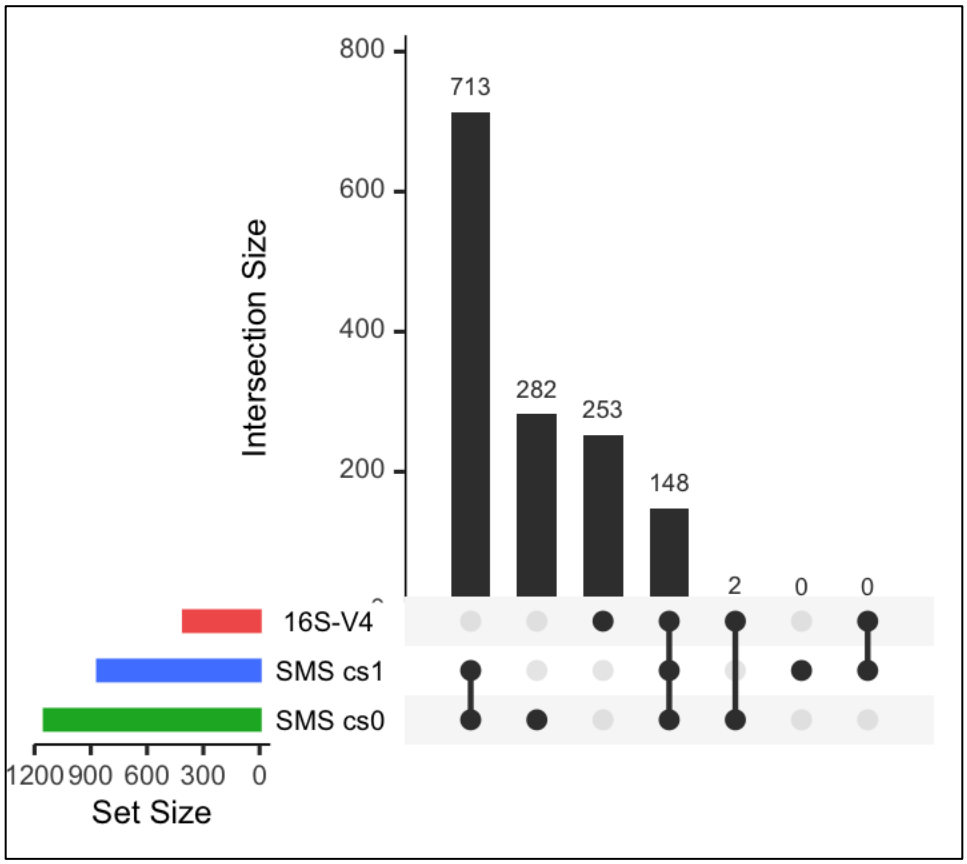


Figure 3.3. Upset plot for genera identified by all 3 methods and <3 methods. Set size (legend) represents the total number of unique genera identified by each method. Intersection size (y-axis) represents the total number of unique genera identified by the corresponding method(s) (x=axis). 16S-V4: 16S rRNA gene sequencing V4 hypervariable region; SMS cs0: Shotgun metagenomic sequencing confidence score 0; SMS cs1: Shotgun metagenomic sequencing confidence score 1.

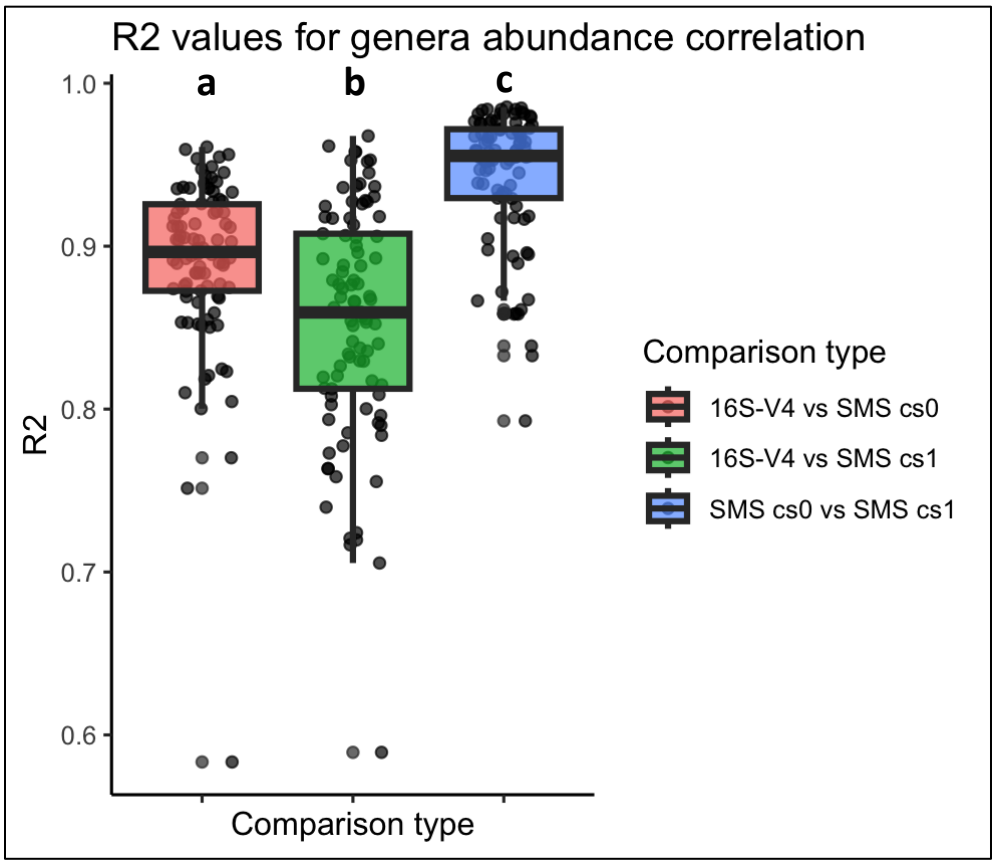


Figure 3.4. Pearson's R^2 values for correlation of genera abundance, comparing the counts obtained by the three methods. 16S-V4: 16S rRNA gene sequencing V4 hypervariable region; SMS cs0: Shotgun metagenomic sequencing confidence score 0; SMS cs1: Shotgun metagenomic sequencing confidence score 1. Boxes represent the 25th to 75th percentile; horizontal line represents the median; and whiskers indicate 1.5× the interquartile range (IQR). Different superscript letters represent values that were significantly different upon statistical testing.

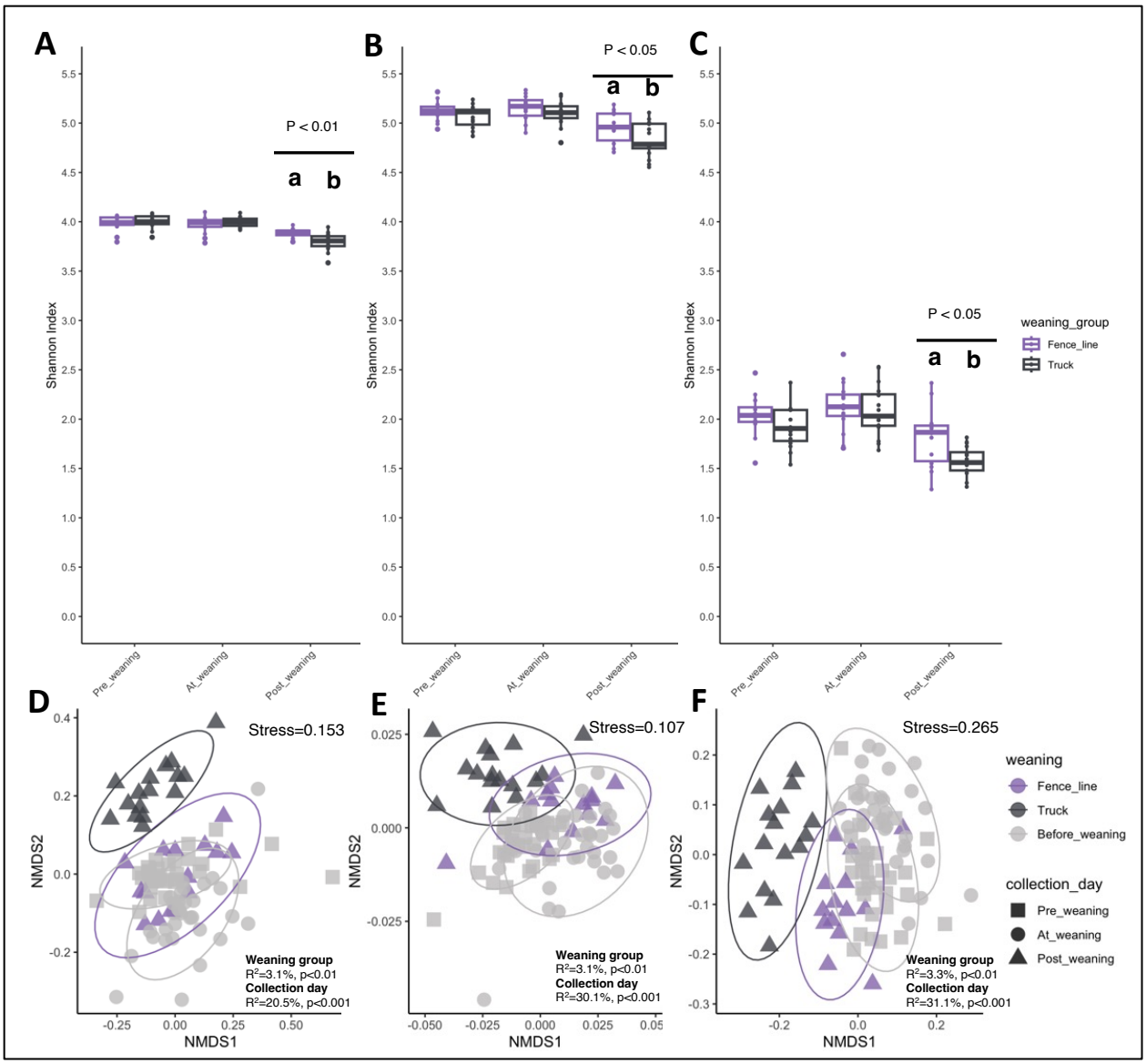


Figure 3.5. Genus-level diversity metrics, stratified by collection day and weaning strategy. Shannon's index (alpha diversity) obtained by A) 16S rRNA gene sequencing, B) Shotgun metagenomic sequencing confidence score 0, and C) confidence score 1. Beta diversity obtained by D) 16S rRNA gene sequencing, E) Shotgun metagenomic sequencing confidence score 0, and F) confidence score 1. Boxes represent the 25th to 75th percentile; horizontal line represents the median; and whiskers indicate 1.5× the interquartile range. *P*-values from Type-III ANOVA. Beta diversity is depicted in a Non-Metric Multidimensional Scaling (NMDS) ordination plot based on Bray–Curtis distances *(P* and R2 values from PERMANOVA testing).

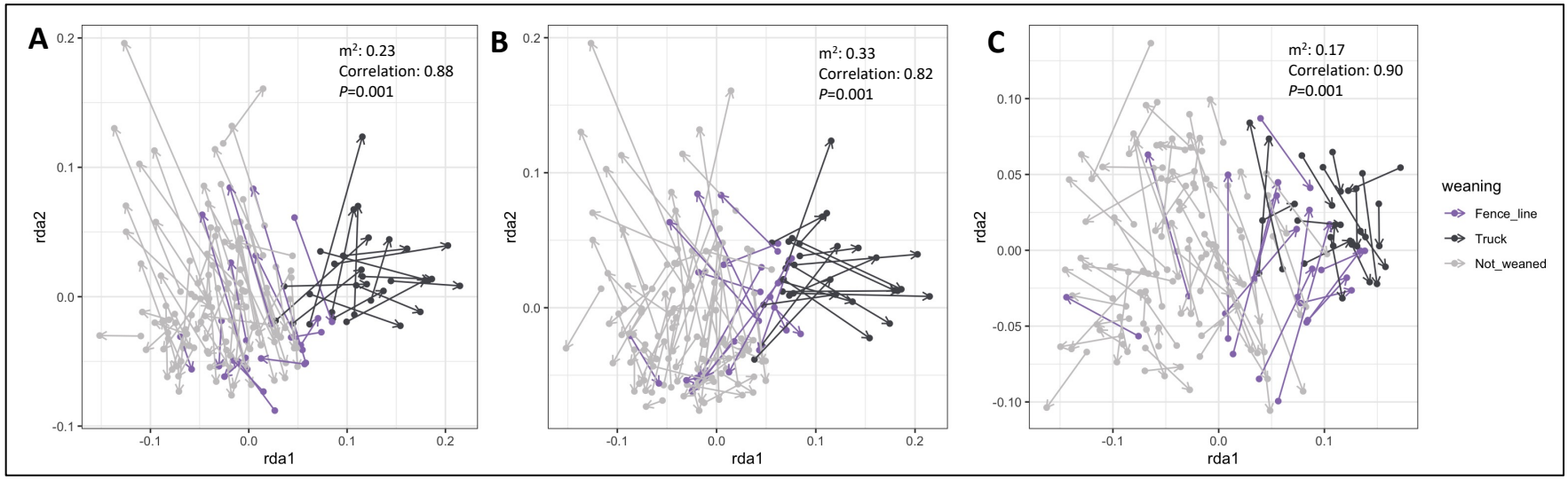


Figure 3.6. Procrustes analysis of genus-level beta diversity plots for the pairwise comparison of A) 16S rRNA gene sequencing vs Shotgun metagenomic sequencing confidence score 0, B) 16S rRNA gene sequencing vs Shotgun metagenomic sequencing confidence score 1, C) Shotgun metagenomic sequencing confidence score 1 vs confidence score 0. *P*-values were calculated by Permutational test of significance of Procrustes analysis with 999 permutations.

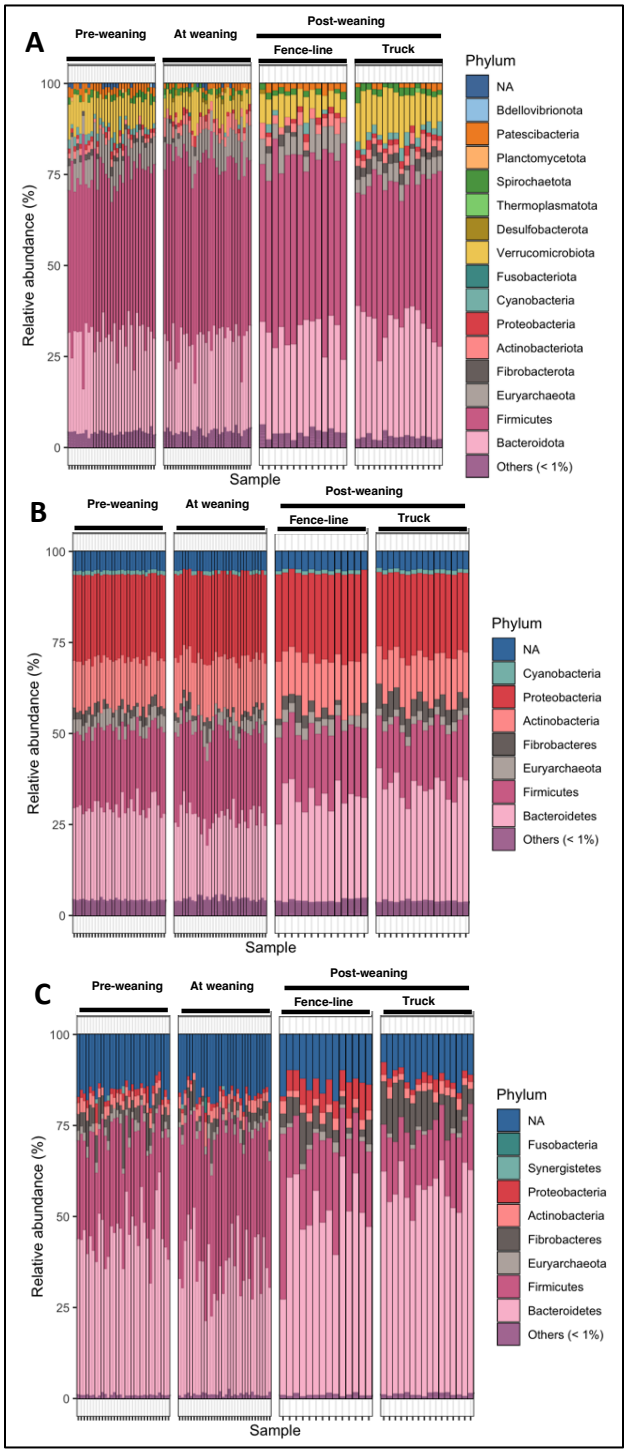
SUPPLEMENTARY MATERIAL

Supplementary table 3.1. Discordant phyla names for the same taxa identified by 16S rRNA gene sequencing (16S-V4) and shotgun metagenomic sequencing (SMS)

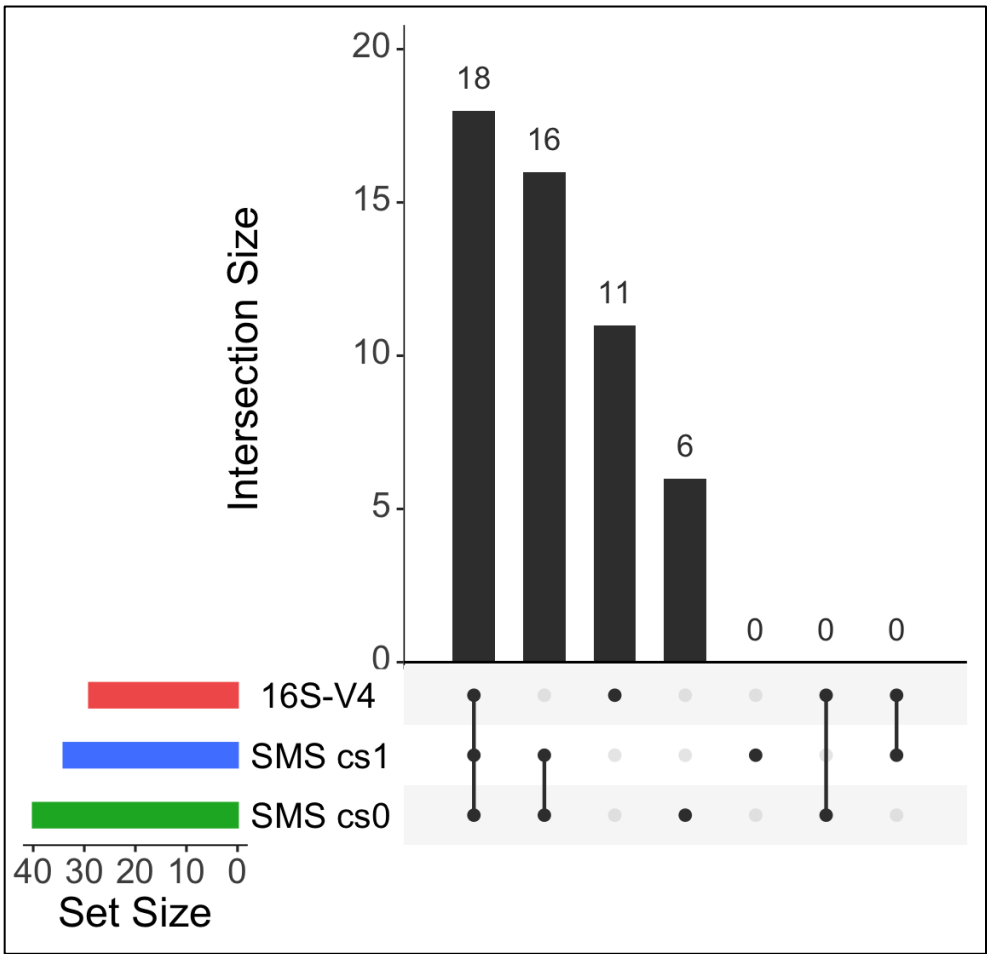
OTU	Kingdom	Phylum	Method
Otu11894	Bacteria	Acidobacteria	SMS
Otu3656	Bacteria	**Acidobacteriota**	16S-V4
Otu08398	Bacteria	**Actinobacteria**	SMS
Otu0035	Bacteria	Actinobacteriota	16S-V4
Otu09386	Bacteria	Armatimonadetes	SMS
Otu1807	Bacteria	**Armatimonadota**	16S-V4
Otu09394	Bacteria	Bacteroidetes	SMS
Otu0007	Bacteria	**Bacteroidota**	16S-V4
Otu11999	Bacteria	Elusimicrobia	SMS
Otu0261	Bacteria	**Elusimicrobiota**	16S-V4
Otu09739	Bacteria	Fibrobacteres	SMS
Otu0025	Bacteria	**Fibrobacterota**	16S-V4
Otu11910	Bacteria	Fusobacteria	SMS
Otu1290	Bacteria	**Fusobacteriota**	16S-V4
Otu09741	Bacteria	Gemmatimonadetes	SMS
Otu7361	Bacteria	**Gemmatimonadota**	16S-V4
Otu11969	Bacteria	Nitrospirae	SMS
Otu3902	Bacteria	**Nitrospirota**	16S-V4
Otu11781	Bacteria	Planctomycetes	SMS
Otu0229	Bacteria	**Planctomycetota**	16S-V4
Otu11830	Bacteria	Spirochaetes	SMS
Otu0118	Bacteria	**Spirochaetota**	16S-V4
Otu11898	Bacteria	Synergistetes	SMS
Otu0537	Bacteria	**Synergistota**	16S-V4
Otu11760	Bacteria	Verrucomicrobia	SMS
Otu0018	Bacteria	**Verrucomicrobiota**	16S-V4

Supplementary table 3.2. Some examples of discordant taxonomy at the phylum and order levels, for the same genus, identified by 16S rRNA gene sequencing (16S-V4) and shotgun metagenomic sequencing (SMS)

OTU	Kingdom	Phylum	Class	Order	Family	Genus	Method
Otu0011	Bacteria	**Bacteroidota**	Bacteroidia	Bacteroidales	Prevotellaceae	**Prevotella**	16S-V4
Otu09394	Bacteria	**Bacteroidetes**	Bacteroidia	Bacteroidales	Prevotellaceae	**Prevotella**	SMS
Otu0025	Bacteria	**Fibrobacterota**	Fibrobacteria	Fibrobacterales	Fibrobacteraceae	**Fibrobacter**	16S-V4
Otu09739	Bacteria	**Fibrobacteres**	Fibrobacteria	Fibrobacterales	Fibrobacteraceae	**Fibrobacter**	SMS
Otu0105	Bacteria	Firmicutes	Clostridia	**Oscillospirales**	Ruminococcaceae	**Ruminococcus**	16S-V4
Otu07736	Bacteria	Firmicutes	Clostridia	**Clostridiales**	Ruminococcaceae	**Ruminococcus**	SMS
Otu0013	Bacteria	Firmicutes	Clostridia	**Lachnospirales**	Lachnospiraceae	**Butyrivibrio**	16S-V4
Otu07712	Bacteria	Firmicutes	Clostridia	**Clostridiales**	Lachnospiraceae	**Butyrivibrio**	SMS



Supplementary Figure 3.1. Relative abundance plots of phylum-level counts obtained by A) 16S rRNA sequencing, B) Shotgun metagenomic cs0, C) Shotgun metagenomic cs1. Phyla with < 1% relative abundance are grouped as "Others". Each bar represents one sample.



Supplementary Figure 3.2. Upset plot for phyla identified by all 3 methods and <3 methods. Set size (legend) represents the total number of unique phyla identified by each method. Intersection size (y-axis) represents the total number of unique phyla identified by the corresponding method(s) (x=axis). 16S-V4: 16S rRNA gene sequencing V4 hypervariable region; SMS cs0: Shotgun metagenomic sequencing confidence score 0; SMS cs1: Shotgun metagenomic sequencing confidence score 1.

GENERAL CONCLUSIONS

Due to its central role in cattle metabolism, the rumen microbiome has been associated with important health and production outcomes in beef cattle. As shown in chapter 01, the scientific literature that describes rumen microbiome development and its association with host traits is less extensive in beef than in dairy cattle. Specifically, the evidence on how the rumen microbiome is impacted by early-life management practices, such as weaning and castration, have not been described yet in beef cattle. At the same time, the methodological approach to study the cattle rumen microbiome is moving from 16S rRNA gene sequencing to shotgun metagenomic sequencing. However, there is still a lack of knowledge on the comparability of these methods. Given that rumen microbiome studies rely on these 2 sequencing methods, it is essential to consider their advantages and disadvantages in order to accurately interpret results. Therefore, the aim of this thesis was to investigate the effect of two early-life common management practices on the rumen metagenome of beef cattle through a comprehensive metagenomic sequencing approach.

In chapter 02, we investigated the effect of four castration timing windows and two weaning strategies on the rumen metagenome of beef calves. After conducting a randomized controlled trial to test these interventions on 32 angus beef calves, we collected rumen microbiome samples at pre-weaning, weaning and post-weaning and analyzed them using shotgun metagenomic sequencing. Our results showed that the rumen microbiome of beef calves experienced significant changes around weaning, with notable shifts in both taxonomic and functional profiles within 48 hours post-weaning, which differed based on weaning strategy. This effect was especially evident in the methanogenic gene content of the rumen microbiome, but not in the rumen resistome, which remained consistently dominated by tetracycline resistance genes across time and intervention groups. Within the context of our study, castration timing had limited long-term effects on the rumen metagenome, although the sampling design may have prevented a definitive conclusion. Further research is needed to explore the short-term effects of castration timing and long-term effects of weaning strategy on the rumen microbiome of beef calves. Despite the limitations of our study, this chapter provided valuable insights into beef calf rumen metagenome dynamics in response to different weaning strategies and castration timing windows. These findings may have direct implications for the management practices that beef producers use to manage calves during the cow-calf stage. Importantly, we

emphasize the need for methodological advancements to improve the metagenomic exploration of the rumen microbiome, paired with a comprehensive phenotypic validation of results from DNA-level rumen microbiome functional profiling.

In chapter 03, we compared the performance of shotgun metagenomic sequencing and 16S rRNA gene hypervariable region V4 sequencing for taxonomic characterization of beef calf rumen microbiomes. We sequenced all the rumen fluid samples from chapter 02 using both methods and compared 3 approaches for taxonomic classification: 16S rRNA gene hypervariable region V4 sequencing, shotgun metagenomic sequencing using Kraken2 with confidence score 0 and with confidence score 1. The results of this study revealed several discrepancies between the methods, which resulted from inherent differences in their technical workflows as well as the many inconsistencies across the databases used for taxonomic classification. Despite these discrepancies, the 3 approaches led to similar ecological inferences about the effect of weaning strategy on the rumen microbiome of beef calves. This important finding is consistent with previous studies of other host-associated microbiomes, and can be explained by a moderate-to-high correlation between the alpha and beta diversity estimates generated by each method, as well as moderate-to-high correlation in the counts of genera that were identified by all 3 methods. We underscored the importance of considering the research goals and available resources while selecting a sequencing method for taxonomic profiling of the rumen microbiome. The results of this chapter will help other researchers seeking accurate methodological approaches to study the rumen microbiome. More studies are warranted to assess other metagenomic taxonomic profilers and the use of other hypervariable regions of the 16S rRNA gene for the taxonomic classification of the rumen microbiome.

Altogether, this thesis contributes new information about the rumen metagenome of beef calves, specifically in relation to early-life management practices such as weaning strategy and castration timing. Moreover, this manuscript provides important technical information about common microbiome sequencing methods and the comparability of the results they generate from rumen samples. This research may have implications for both beef cattle farming and animal microbiome research. Finally, we emphasize the need for phenotypic validation studies and the development of bioinformatic tools and databases in order to continuously improve the robustness of rumen microbiome results.

BIBLIOGRAPHY

1. Greenwood PL. Review: An overview of beef production from pasture and feedlot globally, as demand for beef and the need for sustainable practices increase. Animal. 2021;15:100295.

2. Rotz CA, Asem-Hiablie S, Place S, Thoma G. Environmental footprints of beef cattle production in the United States. Agric Syst. 2019;169:1–13.

3. Murray CF, Fick LJ, Pajor EA, Barkema HW, Jelinski MD, Windeyer MC. Calf management practices and associations with herd-level morbidity and mortality on beef cow-calf operations. animal. 2016;10:468–77.

4. Drouillard JS. Current situation and future trends for beef production in the United States of America — A review. Asian-Australas J Anim Sci. 2018;31:1007–16.

5. Hilton WM. Management of Preconditioned Calves and Impacts of Preconditioning. Vet Clin North Am Food Anim Pract. 2015;31:197–207.

6. Mizrahi I, Wallace RJ, Moraïs S. The rumen microbiome: balancing food security and environmental impacts. Nat Rev Microbiol. 2021;19:553–66.

7. O'Hara E, Neves ALA, Song Y, Guan LL. The Role of the Gut Microbiome in Cattle Production and Health: Driver or Passenger? Annu Rev Anim Biosci. 2020;8:199–220.

8. Clemmons BA, Voy BH, Myer PR. Altering the Gut Microbiome of Cattle: Considerations of Host-Microbiome Interactions for Persistent Microbiome Manipulation. Microb Ecol. 2019;77:523–36.

9. Wallace RJ, Sasson G, Garnsworthy PC, Tapio I, Gregson E, Bani P, et al. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. Sci Adv. 2019;5:eaav8391.

10. Auffret MD, Stewart RD, Dewhurst RJ, Duthie CA, Watson M, Roehe R. Identification of Microbial Genetic Capacities and Potential Mechanisms Within the Rumen Microbiome Explaining Differences in Beef Cattle Feed Efficiency. Front Microbiol. 2020;11:1–16.

11. Newbold CJ, Ramos-Morales E. Review: Ruminal microbiome and microbial metabolome: effects of diet and ruminant host. Animal. 2020;14:s78–86.

12. Ishaq SL, AlZahal O, Walker N, McBride B. An Investigation into Rumen Fungal and Protozoal Diversity in Three Rumen Fractions, during High-Fiber or Grain-Induced Sub-Acute Ruminal Acidosis Conditions, with or without Active Dry Yeast Supplementation. Front Microbiol [Internet]. 2017 [cited 2023 Jan 3];8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.01943

13. Li Z, Wang X, Zhang Y, Yu Z, Zhang T, Dai X, et al. Genomic insights into the phylogeny and biomass-degrading enzymes of rumen ciliates. ISME J. 2022;16:2775–87.

14. Gilbert RA, Kelly WJ, Altermann E, Leahy SC, Minchin C, Ouwerkerk D, et al. Toward Understanding Phage:Host Interactions in the Rumen; Complete Genome Sequences of Lytic Phages Infecting Rumen Bacteria. Front Microbiol [Internet]. 2017 [cited 2023 Jan 3];8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02340

15. Furman O, Shenhav L, Sasson G, Kokou F, Honig H, Jacoby S, et al. Stochasticity constrained by deterministic effects of diet and age drive rumen microbiome assembly dynamics. Nat Commun. 2020;11:1904.

16. Mizrahi I, Jami E. A method to the madness. EMBO Rep. 2021;22:2–5.

17. Guzman CE, Bereza-Malcolm LT, Groef BD, Franks AE. Presence of Selected Methanogens, Fibrolytic Bacteria, and Proteobacteria in the Gastrointestinal Tract of Neonatal Dairy Calves from Birth to 72 Hours. PLOS ONE. 2015;10:e0133048.

18. O'Hara E, Kenny DA, McGovern E, Byrne CJ, McCabe MS, Guan LL, et al. Investigating temporal microbial dynamics in the rumen of beef calves raised on two farms during early life. FEMS Microbiol Ecol. 2020;96:fiz203.

19. Lourenco JM, Krause TR, Welch CB, Callaway TR, Pringle TD. Longitudinal Changes of the Ruminal Microbiota in Angus Beef Steers. Anim Open Access J MDPI. 2022;12:1066.

20. Lourenco JM, Kieran TJ, Seidel DS, Glenn TC, Da Silveira MF, Callaway TR, et al. Comparison of the ruminal and fecal microbiotas in beef calves supplemented or not with concentrate. PLoS ONE. 2020;15:1–17.

21. Lourenco JM, Callaway TR, Kieran TJ, Glenn TC, McCann JC, Stewart RL. Analysis of the Rumen Microbiota of Beef Calves Supplemented During the Suckling Phase. Front Microbiol [Internet]. 2019 [cited 2023 May 17];10. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01131

22. Omontese BO, Sharma AK, Davison S, Jacobson E, DiConstanzo A, Webb MJ, et al. Microbiome network traits in the rumen predict average daily gain in beef cattle under different backgrounding systems. Anim Microbiome [Internet]. 2022;4. Available from: https://doi.org/10.1186/s42523-022-00175-y

23. Li F, Zhou M, Ominski K, Guan LL. Does the rumen microbiome play a role in feed efficiency of beef cattle?1. J Anim Sci. 2016;94:44–8.

24. Tapio I, Snelling TJ, Strozzi F, Wallace RJ. The ruminal microbiome associated with methane emissions from ruminant livestock. J Anim Sci Biotechnol. 2017;8:7.

25. Jami E, White BA, Mizrahi I. Potential Role of the Bovine Rumen Microbiome in Modulating Milk Composition and Feed Efficiency. PLoS ONE. 2014;9:e85423.

26. Stergiadis S, Cabeza-Luna I, Mora-Ortiz M, Stewart RD, Dewhurst RJ, Humphries DJ, et al. Unravelling the Role of Rumen Microbial Communities, Genes, and Activities on Milk Fatty Acid Profile Using a Combination of Omics Approaches. Front Microbiol. 2020;11:590441.

27. Nkrumah JD, Okine EK, Mathison GW, Schmid K, Li C, Basarab JA, 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. 2006;84:145–53.

28. Elolimy AA, Abdelmegeid MK, McCann JC, Shike DW, Loor JJ. Residual feed intake in beef cattle and its association with carcass traits, ruminal solid-fraction bacteria, and epithelium gene expression. J Anim Sci Biotechnol. 2018;9:67.

29. Abbas W, Howard JT, Paz HA, Hales KE, Wells JE, Kuehn LA, et al. Influence of host genetics in shaping the rumen bacterial community in beef cattle. Sci Rep. 2020;10:1–14.

30. Auffret MD, Dewhurst RJ, Duthie C-A, Rooke JA, John Wallace R, Freeman TC, et al. The rumen microbiome as a reservoir of antimicrobial resistance and pathogenicity genes is directly affected by diet in beef cattle. Microbiome. 2017;5:159.

31. Wallace RJ, Rooke JA, McKain N, Duthie CA, Hyslop JJ, Ross DW, et al. The rumen microbial metagenome associated with high methane production in cattle. BMC Genomics. 2015;16:1–14.

32. Danielsson R, Dicksved J, Sun L, Gonda H, Müller B, Schnürer A, et al. Methane Production in Dairy Cows Correlates with Rumen Methanogenic and Bacterial Community Structure. Front Microbiol [Internet]. 2017 [cited 2023 Jan 3];8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.00226

33. Auffret MD, Stewart R, Dewhurst RJ, Duthie C-A, Rooke JA, Wallace RJ, et al. Identification, Comparison, and Validation of Robust Rumen Microbial Biomarkers for Methane Emissions Using Diverse Bos Taurus Breeds and Basal Diets. Front Microbiol. 2018;8:2642.

34. Pereira AM, de Lurdes Nunes Enes Dapkevicius M, Borba AES. Alternative pathways for hydrogen sink originated from the ruminal fermentation of carbohydrates: Which microorganisms are involved in lowering methane emission? Anim Microbiome. 2022;4:5.

35. Kim M, Park T, Jeong JY, Baek Y, Lee H-J. Association between Rumen Microbiota and Marbling Score in Korean Native Beef Cattle. Animals. 2020;10:712.

36. Krause TR, Lourenco JM, Welch CB, Rothrock MJ, Callaway TR, Pringle TD. The relationship between the rumen microbiome and carcass merit in Angus steers. J Anim Sci. 2020;98:skaa287.

37. Fang S, Xiong X, Su Y, Huang L, Chen C. 16S rRNA gene-based association study identified microbial taxa associated with pork intramuscular fat content in feces and cecum lumen. BMC Microbiol. 2017;17:162.

38. Martínez-Álvaro M, Mattock J, Auffret M, Weng Z, Duthie C-A, Dewhurst RJ, et al. Microbiome-driven breeding strategy potentially improves beef fatty acid profile benefiting human health and reduces methane emissions. Microbiome. 2022;10:166.

39. Li F, Li C, Chen Y, Liu J, Zhang C, Irving B, et al. Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. Microbiome. 2019;7:92.

40. Ma T, Zaheer R, McAllister TA, Guo W, Li F, Tu Y, et al. Expressions of resistome is linked to the key functions and stability of active rumen microbiome. Anim Microbiome. 2022;4:38.

41. Nearing JT, Comeau AM, Langille MGI. Identifying biases and their potential solutions in human microbiome studies. Microbiome. 2021;9:113.

42. De Menezes AB, Lewis E, O'Donovan M, O'Neill BF, Clipson N, Doyle EM. Microbiome analysis of dairy cows fed pasture or total mixed ration diets: Microbiome analysis of rumen contents from dairy cows. FEMS Microbiol Ecol. 2011;78:256–65.

43. Weinroth MD, Belk AD, Dean C, Noyes N, Dittoe DK, Rothrock MJ, et al. Considerations and best practices in animal science 16S ribosomal RNA gene sequencing microbiome studies. J Anim Sci. 2022;100:skab346.

44. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017;35:833–44.

45. Anderson CL, Fernando SC. Insights into rumen microbial biosynthetic gene cluster diversity through genome-resolved metagenomics. Commun Biol. 2021;4:1–12.

46. Smith RH, Glendinning L, Walker AW, Watson M. Investigating the impact of database choice on the accuracy of metagenomic read classification for the rumen microbiome. 2022;1–39.

47. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, et al. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nat Commun. 2018;9:1–11.

48. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. Nat Biotechnol. 2019;37:953–61.

49. Zhou M, O'Hara E, Tang S, Chen Y, Walpole ME, Górka P, et al. Accessing Dietary Effects on the Rumen Microbiome: Different Sequencing Methods Tell Different Stories. Vet Sci. 2021;8:138.

50. Roehe R, Dewhurst RJ, Duthie C-A, Rooke JA, McKain N, Ross DW, 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. 2016;12:e1005846.

51. Liu J, Bai Y, Liu F, Kohn RA, Tadesse DA, Sarria S, et al. Rumen Microbial Predictors for Short-Chain Fatty Acid Levels and the Grass-Fed Regimen in Angus Cattle. Animals. 2022;12:2995.

52. Meale SJ, Li S, Azevedo P, Derakhshani H, Plaizier JC, Khafipour E, et al. Development of Ruminal and Fecal Microbiomes Are Affected by Weaning But Not Weaning Strategy in Dairy Calves. Front Microbiol [Internet]. 2016 [cited 2022 Sep 7];7. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2016.00582

53. Meale SJ, Li SC, Azevedo P, Derakhshani H, DeVries TJ, Plaizier JC, et al. Weaning age influences the severity of gastrointestinal microbiome shifts in dairy calves. Sci Rep. 2017;7:198.

54. Amin N, Schwarzkopf S, Kinoshita A, Tröscher-Mußotter J, Dänicke S, Camarinha-Silva A, et al. Evolution of rumen and oral microbiota in calves is influenced by age and time of weaning. Anim Microbiome. 2021;3.

55. Woodruff KL, Hummel GL, Austin KJ, Lake SL, Cunningham-Hollinger HC. Calf rumen microbiome from birth to weaning and shared microbial properties to the maternal rumen microbiome. J Anim Sci. 2022;100:skac264.

56. Baek YC, Choi H, Jeong J-Y, Lee SD, Kim MJ, Lee S, et al. The impact of short-term acute heat stress on the rumen microbiome of Hanwoo steers. J Anim Sci Technol. 2020;62:208–17.

57. Mir RA, Kleinhenz MD, Coetzee JF, Allen HK, Kudva IT. Fecal microbiota changes associated with dehorning and castration stress primarily affects light-weight dairy calves. PLoS ONE. 2019;14:1–14.

58. Correia Sales GF, Carvalho BF, Schwan RF, de Figueiredo Vilela L, Moreno Meneses JA, Gionbelli MP, et al. Heat stress influence the microbiota and organic acids concentration in beef cattle rumen. J Therm Biol. 2021;97:102897.

59. Paz HA, Hales KE, Wells JE, Kuehn LA, Freetly HC, Berry ED, et al. Rumen bacterial community structure impacts feed efficiency in beef cattle. J Anim Sci. 2018;96:1045–58.

60. Gerber PJ, Mottet A, Opio CI, Falcucci A, Teillard F. Environmental impacts of beef production: Review of challenges and perspectives for durability. Meat Sci. 2015;109:2–12.

61. Martínez-Álvaro M, Auffret MD, Stewart RD, Dewhurst RJ, Duthie C-A, Rooke JA, et al. Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine. Front Microbiol [Internet]. 2020 [cited 2022 Oct 29];11. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00659

62. Tucker CB, Coetzee JF, Stookey JM, Thomson DU, Grandin T, Schwartzkopf-Genswein KS. Beef cattle welfare in the USA: identification of priorities for future research. Anim Health Res Rev. 2015;16:107–24.

63. Chen Y, Arsenault R, Napper S, Griebel P. Models and Methods to Investigate Acute Stress Responses in Cattle. Anim Open Access J MDPI. 2015;5:1268–95.

64. Roberts SL, Powell JG, Hughes HD, Richeson JT. Effect of castration method and analgesia on inflammation, behavior, growth performance, and carcass traits in feedlot cattle. J Anim Sci. 2018;96:66–75.

65. Lynch E, McGee M, Earley B. Weaning management of beef calves with implications for animal health and welfare. J Appl Anim Res. 2019;47:167–75.

66. Canozzi MEA, Mederos A, Turner S, Manteca X, McManus C, Menegassi SRO, et al. Dehorning and welfare indicators in beef cattle: a meta-analysis. Anim Prod Sci. 2018;59:801–14.

67. Steen RWJ, Lavery NP, Kilpatrick DJ, Porter MG. Effects of pasture and high-concentrate diets on the performance of beef cattle, carcass composition at equal growth rates, and the fatty acid composition of beef. N Z J Agric Res. 2003;46:69–81.

68. Loor JJ, Elolimy AA, McCann JC. Dietary impacts on rumen microbiota in beef and dairy production. Anim Front. 2016;6:22–9.

69. Ogunade IM, Lay J, Andries K, McManus CJ, Bebe F. Effects of live yeast on differential genetic and functional attributes of rumen microbiota in beef cattle. J Anim Sci Biotechnol. 2019;10:68.

70. Zhou K, Bao Y, Zhao G. Effects of dietary crude protein and tannic acid on rumen fermentation, rumen microbiota and nutrient digestion in beef cattle. Arch Anim Nutr. 2019;73:30–43.

71. Woodruff KL, Hummel GL, Austin KJ, Smith TL, Cunningham-Hollinger HC. Influence of the maternal rumen microbiome on development of the calf meconium and rumen microbiome. Transl Anim Sci. 2020;4:S169–73.

72. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. Nat Biotechnol. 2018;36:359–67.

73. Qian L, Yu X, Zhou J, Gu H, Ding J, Peng Y, et al. MCycDB: A curated database for comprehensively profiling methane cycling processes of environmental microbiomes. Mol Ecol Resour. 2022;22:1803–23.

74. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, et al. MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res. 2020;48:D561–9.

75. Whon TW, Kim HS, Shin N, Jung ES, Tak EJ, Sung H, et al. Male castration increases adiposity via small intestinal microbial alterations. EMBO Rep. 2021;22:1–18.

76. Li Z, Shi J, Lei Y, Wu J, Zhang R, Zhang X, et al. Castration alters the cecal microbiota and inhibits growth in Holstein cattle. J Anim Sci. 2022;100:skac367.

77. Zhang Z, Yang L, He Y, Luo X, Zhao S, Jia X. Composition of Fecal Microbiota in Grazing and Feedlot Angus Beef Cattle. Anim Open Access J MDPI. 2021;11:3167.

78. Bailoni L, Carraro L, Cardin M, Cardazzo B. Active Rumen Bacterial and Protozoal Communities Revealed by RNA-Based Amplicon Sequencing on Dairy Cows Fed Different Diets at Three Physiological Stages. Microorganisms. 2021;9:754.

79. Rey M, Enjalbert F, Combes S, Cauquil L, Bouchez O, Monteils V. Establishment of ruminal bacterial community in dairy calves from birth to weaning is sequential. J Appl Microbiol. 2014;116:245–57.

80. Wang S, Chai J, Zhao G, Zhang N, Cui K, Bi Y, et al. The Temporal Dynamics of Rumen Microbiota in Early Weaned Lambs. Microorganisms. 2022;10:144.

81. Zhuang Y, Chai J, Cui K, Bi Y, Diao Q, Huang W, et al. Longitudinal Investigation of the Gut Microbiota in Goat Kids from Birth to Postweaning. Microorganisms. 2020;8:1111.

82. Enríquez D, Hötzel MJ, Ungerfeld R. Minimising the stress of weaning of beef calves: a review. Acta Vet Scand. 2011;53:28.

83. Taylor JD, Gilliam JN, Mourer G, Stansberry C. Comparison of effects of four weaning methods on health and performance of beef calves. animal. 2020;14:161–70.

84. Guevarra RB, Hong SH, Cho JH, Kim B-R, Shin J, Lee JH, et al. The dynamics of the piglet gut microbiome during the weaning transition in association with health and nutrition. J Anim Sci Biotechnol. 2018;9:54.

85. Mach N, Foury A, Kittelmann S, Reigner F, Moroldo M, Ballester M, et al. The Effects of Weaning Methods on Gut Microbiota Composition and Horse Physiology. Front Physiol [Internet]. 2017 [cited 2023 Jan 14];8. Available from: https://www.frontiersin.org/articles/10.3389/fphys.2017.00535

86. Wang O, McAllister TA, Plastow G, Stanford K, Selinger B, Guan LL. Interactions of the Hindgut Mucosa-Associated Microbiome with Its Host Regulate Shedding of Escherichia coli O157:H7 by Cattle. Appl Environ Microbiol. 2017;84:e01738-17.

87. Chen S, Wang J, Peng D, Li G, Chen J, Gu X. Exposure to heat-stress environment affects the physiology, circulation levels of cytokines, and microbiome in dairy cows. Sci Rep. 2018;8:14606.

88. Deng L, He C, Zhou Y, Xu L, Xiong H. Ground transport stress affects bacteria in the rumen of beef cattle: A real-time PCR analysis. Anim Sci J Nihon Chikusan Gakkaiho. 2017;88:790–7.

89. Beaber JW, Hochhut B, Waldor MK. SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature. 2004;427:72–4.

90. Samuels AN, Roggiani M, Zhu J, Goulian M, Kohli RM. The SOS Response Mediates Sustained Colonization of the Mammalian Gut. Infect Immun. 2019;87:e00711-18.

91. Weiss GA, Hennet T. Mechanisms and consequences of intestinal dysbiosis. Cell Mol Life Sci. 2017;74:2959–77.

92. Betancur-Murillo CL, Aguilar-Marín SB, Jovel J. Prevotella: A Key Player in Ruminal Metabolism. Microorganisms. 2022;11:1.

93. Weimer PJ. Degradation of Cellulose and Hemicellulose by Ruminal Microorganisms. Microorganisms. 2022;10:2345.

94. Foster JA, Rinaman L, Cryan JF. Stress & the gut-brain axis: Regulation by the microbiome. Neurobiol Stress. 2017;7:124–36.

95. Mizrahi I, Wallace RJ, Moraïs S. The rumen microbiome: balancing food security and environmental impacts. Nat Rev Microbiol. 2021;19:553–66.

96. Cunha CS, Marcondes MI, Veloso CM, Mantovani HC, Pereira LGR, Tomich TR, et al. Compositional and structural dynamics of the ruminal microbiota in dairy heifers and its relationship to methane production. J Sci Food Agric. 2019;99:210–8.

97. Ramayo-Caldas Y, Zingaretti L, Popova M, Estellé J, Bernard A, Pons N, et al. Identification of rumen microbial biomarkers linked to methane emission in Holstein dairy cows. J Anim Breed Genet Z Tierzuchtung Zuchtungsbiologie. 2020;137:49–59.

98. Cardinale S, Kadarmideen HN. Host Genome–Metagenome Analyses Using Combinatorial Network Methods Reveal Key Metagenomic and Host Genetic Features for Methane Emission and Feed Efficiency in Cattle. Front Genet [Internet]. 2022 [cited 2022 Sep 7];13. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2022.795717

99. Brown Kav A, Sasson G, Jami E, Doron-Faigenboim A, Benhar I, Mizrahi I. Insights into the bovine rumen plasmidome. Proc Natl Acad Sci U S A. 2012;109:5452–7.

100. Sabino YNV, Santana MF, Oyama LB, Santos FG, Moreira AJS, Huws SA, et al. Characterization of antibiotic resistance genes in the species of the rumen microbiota. Nat Commun. 2019;10:5252.

101. López-Catalina A, Atxaerandio R, García-Rodríguez A, Goiri I, Gutierrez-Rivas M, Jiménez-Montero JA, et al. Characterisation of the rumen resistome in Spanish dairy cattle. Anim Microbiome. 2021;3:63.

102. Xue M-Y, Xie Y-Y, Zhong Y-F, Liu J-X, Guan LL, Sun H-Z. Ruminal resistome of dairy cattle is individualized and the resistotypes are associated with milking traits. Anim Microbiome. 2021;3:18.

103. Noyes NR, Yang X, Linke LM, Magnuson RJ, Cook SR, Zaheer R, et al. Characterization of the resistome in manure, soil and wastewater from dairy and beef production systems. Sci Rep. 2016;6:24645.

104. Noyes NR, Yang X, Linke LM, Magnuson RJ, Dettenwanger A, Cook S, et al. Resistome diversity in cattle and the environment decreases during beef production. Cooper B, editor. eLife. 2016;5:e13195.

105. Noyes NR, Weinroth ME, Parker JK, Dean CJ, Lakin SM, Raymond RA, et al. Enrichment allows identification of diverse, rare elements in metagenomic resistome-virulome sequencing. Microbiome. 2017;5:142.

106. Liu J, Taft DH, Maldonado-Gomez MX, Johnson D, Treiber ML, Lemay DG, et al. The fecal resistome of dairy cattle is associated with diet during nursing. Nat Commun. 2019;10:4406.

107. Pinnell LJ, Kuiper G, Huebner KL, Doster E, Parker JK, Alekozai N, et al. More than an anthropogenic phenomenon: Antimicrobial resistance in ungulates from natural and agricultural environments. Sci Total Environ. 2023;858:159789.

108. Zamkovaya T, Foster JS, de Crécy-Lagard V, Conesa A. A network approach to elucidate and prioritize microbial dark matter in microbial communities. ISME J. 2021;15:228–44.

109. Azua-Bustos A, Fairén AG, González-Silva C, Prieto-Ballesteros O, Carrizo D, Sánchez-García L, et al. Dark microbiome and extremely low organics in Atacama fossil delta unveil Mars life detection limits. Nat Commun. 2023;14:808.

110. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. mSystems. 2018;3:e00055-18.

111. Regalado J, Lundberg DS, Deusch O, Kersten S, Karasov T, Poersch K, et al. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe–microbe interaction networks in plant leaves. ISME J. 2020;14:2116–30.

112. Beck KL, Haiminen N, Chambliss D, Edlund S, Kunitomi M, Huang BC, et al. Monitoring the microbiome for food safety and quality using deep shotgun sequencing. Npj Sci Food. 2021;5:3.

113. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15.

114. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:1–13.

115. Escudeiro P, Henry CS, Dias RPM. Functional characterization of prokaryotic dark matter: the road so far and what lies ahead. Curr Res Microb Sci. 2022;3:100159.

116. Zehavi T, Probst M, Mizrahi I. Insights Into Culturomics of the Rumen Microbiome. Front Microbiol [Internet]. 2018 [cited 2022 Dec 23];9. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01999

117. Yan M, Pratama AA, Li Z, Jiang Y, Sullivan MB, Yu Z. Unraveling the viral dark matter of the rumen microbiome with a new global virome database [Internet]. bioRxiv; 2022 [cited 2023 Apr 23]. p. 2022.11.30.518432. Available from: https://www.biorxiv.org/content/10.1101/2022.11.30.518432v1

118. Brown Kav A, Rozov R, Bogumil D, Sørensen SJ, Hansen LH, Benhar I, et al. Unravelling plasmidome distribution and interaction with its hosting microbiome. Environ Microbiol. 2020;22:32–44.

119. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

120. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

121. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

122. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:1–13.

123. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30:614–20.

124. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.

125. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLOS ONE. 2016;11:e0163962.

126. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE. 2013;8:e61217.

127. Paulson JN, Colin Stine O, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods. 2013;10:1200–2.

128. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003;14:927–30.

129. Weinroth MD, Belk AD, Dean C, Noyes N, Dittoe DK, Rothrock MJ, et al. Considerations and best practices in animal science 16S ribosomal RNA gene sequencing microbiome studies. J Anim Sci. 2022;100:1–18.

130. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. New York, NY: Springer; 2009 [cited 2023 Mar 13]. Available from: https://link.springer.com/10.1007/978-0-387-98141-3

131. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

132. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 2015;67:1–48.

133. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. J Stat Softw. 2017;82:1–26.

134. Lenth RV, Bolker B, Buerkner P, Giné-Vázquez I, Herve M, Jung M, et al. emmeans: Estimated Marginal Means, aka Least-Squares Means [Internet]. 2023 [cited 2023 Mar 13]. Available from: https://CRAN.R-project.org/package=emmeans

135. Fox J, Weisberg S, Price B, Adler D, Bates D, Baud-Bovy G, et al. car: Companion to Applied Regression [Internet]. 2022 [cited 2023 Mar 13]. Available from: https://CRAN.R-project.org/package=car

136. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. Nat Rev Microbiol. 2021;19:55–71.

137. Chen S, Luo S, Yan C. Gut Microbiota Implications for Health and Welfare in Farm Animals: A Review. Anim Open Access J MDPI. 2021;12:93.

138. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol. 2017;15:579–90.

139. Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform. 2021;22:178–93.

140. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.

141. Peterson D, Bonham KS, Rowland S, Pattanayak CW, RESONANCE Consortium, Klepac-Ceraj V, et al. Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes. Front Microbiol [Internet]. 2021 [cited 2023 Jan 29];12. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.670336

142. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11:2639–43.

143. Hassler HB, Probert B, Moore C, Lawson E, Jackson RW, Russell BT, et al. Phylogenies of the 16S rRNA gene and its hypervariable regions lack concordance with core genome phylogenies. Microbiome. 2022;10:104.

144. Bukin YS, Galachyants YP, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI. The effect of 16s rRNA region choice on bacterial community metabarcoding results. Sci Data. 2019;6:1–14.

145. Toole DR, Zhao J, Martens-Habbena W, Strauss SL. Bacterial functional prediction tools detect but underestimate metabolic diversity compared to shotgun metagenomics in southwest Florida soils. Appl Soil Ecol. 2021;168:104129.

146. Djemiel C, Maron P-A, Terrat S, Dequiedt S, Cottin A, Ranjard L. Inferring microbiota functions from taxonomic genes: a review. GigaScience. 2022;11:giab090.

147. Heidrich V, Beule L. Are short-read amplicons suitable for the prediction of microbiome functional potential? A critical perspective. iMeta. 2022;1:e38.

148. Sun Z, Huang S, Zhang M, Zhu Q, Haiminen N, Carrieri AP, et al. Challenges in Benchmarking Metagenomic Profilers. Nat Methods. 2021;18:618–26.

149. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking Metagenomics Tools for Taxonomic Classification. Cell. 2019;178:779–94.

150. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015;12:902–3.

151. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7:11257.

152. Scholz M, Ward DV, Pasolli E, Tolio T, Zolfo M, Asnicar F, et al. Strain-level microbial epidemiology and population genomics from shotgun metagenomics. Nat Methods. 2016;13:435–8.

153. Usyk M, Peters BA, Karthikeyan S, McDonald D, Sollecito CC, Vazquez-Baeza Y, et al. Comprehensive Evaluation of Shotgun Metagenomics, Amplicon Sequencing and Harmonization of Said Platforms for Epidemiological Studies Using the Large Multi-Center HCHS/SOL Cohort [Internet]. Rochester, NY; 2022 [cited 2023 Jan 29]. Available from: https://papers.ssrn.com/abstract=4197397

154. Brumfield KD, Huq A, Colwell RR, Olds JL, Leddy MB. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. PLOS ONE. 2020;15:e0228899.

155. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. Sci Rep. 2021;11:3030.

156. Huws SA, Creevey CJ, Oyama LB, Mizrahi I, Denman SE, Popova M, et al. Addressing global ruminant agricultural challenges through understanding the rumen microbiome: Past, present, and future. Front Microbiol. 2018;9:1–33.

157. Myer PR, Smith TPL, Wells JE, Kuehn LA, Freetly HC. Rumen Microbiome from Steers Differing in Feed Efficiency. PLOS ONE. 2015;10:e0129174.

158. Xie Y, Sun H, Xue M, Liu J. Metagenomics reveals differences in microbial composition and metabolic functions in the rumen of dairy cows with different residual feed intake. Anim Microbiome. 2022;4:19.

159. Zuo W, Wang B, Bai X, Luan Y, Fan Y, Michail S, et al. 16S rRNA and metagenomic shotgun sequencing data revealed consistent patterns of gut microbiome signature in pediatric ulcerative colitis. Sci Rep. 2022;12:6421.

160. Stothart MR, McLoughlin PD, Poissant J. Shallow shotgun sequencing of the microbiome recapitulates 16S amplicon results and provides functional insights. Mol Ecol Resour. 2023;23:549–64.

161. Bell KL, Petit III RA, Cutler A, Dobbs EK, Macpherson JM, Read TD, et al. Comparing whole-genome shotgun sequencing and DNA metabarcoding approaches for species identification and quantification of pollen species mixtures. Ecol Evol. 2021;11:16082–98.

162. Tessler M, Neumann JS, Afshinnekoo E, Pineda M, Hersch R, Velho LFM, et al. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Sci Rep. 2017;7:6589.

163. Balvočiūtė M, Huson DH. SILVA, RDP, Greengenes, NCBI and OTT — how do these taxonomies compare? BMC Genomics. 2017;18:114.

164. Louca S, Doebeli M, Parfrey LW. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome. 2018;6:41.

165. Sze MA, Schloss PD. The Impact of DNA Polymerase and Number of Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data. mSphere. 2019;4:e00163-19.

166. Silverman JD, Bloom RJ, Jiang S, Durand HK, Dallow E, Mukherjee S, et al. Measuring and mitigating PCR bias in microbiota datasets. PLOS Comput Biol. 2021;17:e1009113.

167. Chiarello M, McCauley M, Villéger S, Jackson CR. Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. PLOS ONE. 2022;17:e0264443.

168. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun. 2019;10:1–11.

169. Matsuo Y, Komiya S, Yasumizu Y, Yasuoka Y, Mizushima K, Takagi T, et al. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. BMC Microbiol. 2021;21:35.

170. Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. Proc Natl Acad Sci. 2015;112:14024–9.

171. Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn L-J, et al. Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. Front Microbiol [Internet]. 2019 [cited 2023 May 4];10. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01277

172. McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing experiments. Turnbaugh P, Garrett WS, Turnbaugh P, Quince C, Gibbons S, editors. eLife. 2019;8:e46923.

173. Han D, Gao P, Li R, Tan P, Xie J, Zhang R, et al. Multicenter assessment of microbial community profiling using 16S rRNA gene sequencing and shotgun metagenomic sequencing. J Adv Res. 2020;26:111–21.

174. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15.

175. Park T, Cersosimo LM, Radloff W, Zanton GI, Li W. The rumen liquid metatranscriptome of post-weaned dairy calves differed by pre-weaning ruminal administration of differentially-enriched, rumen-derived inocula. Anim Microbiome. 2022;4:4.

176. Wright RJ, Comeau AM, Langille MGI. From defaults to databases: parameter and database choice dramatically impact the performance of metagenomic taxonomic classification tools [Internet]. bioRxiv; 2022 [cited 2023 Feb 11]. p. 2022.04.27.489753. Available from: https://www.biorxiv.org/content/10.1101/2022.04.27.489753v1

177. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10–2.

178. Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nat Biotechnol. 2016;34:942–9.

179. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol. 2007;73:5261–7.

180. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6:226.

181. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

182. Xu S, Zhan L, Tang W, Wang Q, Dai Z, Zhou L, et al. MicrobiotaProcess: A comprehensive R package for deep mining microbiome. The Innovation. 2023;4:100388.

183. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinforma Oxf Engl. 2017;33:2938–40.

184. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P, O'Hara B, et al. Vegan: Community Ecology Package. R Package Version 22-1. 2015;2:1–2.

185. Cailliez F. The analytical solution of the additive constant problem. Psychometrika. 1983;48:305–8.