Profiling true microbial community of newborn beef calves using low microbial biomass samples

by

Ranga Nakandalage Don

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Animal Science

Department of Agricultural, Food and Nutritional Science

University of Alberta

Abstract

Amplicon sequencing (16S rRNA gene sequencing) is widely used to profile host-associated microbial communities. Rapid advancements in sequencing and user-friendly bioinformatics platforms have improved the knowledge of microbial community composition. However, profiling samples containing low microbial biomass (biological samples containing limited microbial materials) using amplicon sequencing is challenging, as the presence of low levels of microbial genetic materials in samples leads to the generation of higher levels of artificial sequences during the sequencing process. The use of appropriate approaches to control contaminations during sample processing and sequencing, and the optimization of bioinformatics pipelines determine the accuracy of next-generation sequencing (NGS) based microbial profiling. QIIME2 is one of the most used bioinformatics pipelines that allows users to perform quality filtering, classification, community analysis, visualization, and statistical analysis through one open-source software package. Denoising in QIIME2 is one of the important plugins for quality filtering, which should be handled carefully to generate credible outcomes from microbial community profiling. The first study (Chapter 2) of this thesis aimed to optimize the denoising parameters to increase the accuracy of microbial community data analysis when using low-microbial biomass samples. This study used primers targeting the V1V3 region of the 16S rRNA gene to profile the fecal microbial communities of newborn beef calves sampled using swabs and data were analyzed using QIIME2 with customized quality filtering steps to remove environmental contaminations and to filter out low abundant amplicon sequencing variants (ASVs). Use of optimized (truncation: forward – 294; reverse – 241; median quality score - ≥25) denoising parameters increased the percentage of merged read (default – 1%; optimized – 45%), and the number of samples used for downstream analysis compared to default approach in QIIME 2, which is based on trimming reads based on mean quality score (truncation: forward – 281; reverse – 207; mean quality score - ≥25). Moreover, the optimization of denoising parameters increased microbial diversity and classified taxa. Our study confirmed that optimization of denoising parameters enhances the accuracy of outcomes and interpretation of host-associated microbial community compositions compared to default denoising parameters, especially when the biological sample contains low microbial biomass. Our findings revealed that the default settings of the bioinformatics tools might not be suitable for all microbial analyses. Customizing parameters in bioinformatics pipelines need to be considered to obtain credible outcomes in microbial community assessments. The second study (chapter 3) compared amplicon sequencing-based microbial profiles

generated by different genetic materials (DNA vs. RNA) and hypervariable regions of the 16S rRNA gene (V1V3 vs. V3V4). Rectal and oral swabs (n=40) were collected from 20 newborn beef calves and used to extract DNA and RNA. Both DNA- and RNA-based amplicon sequencing were performed by targeting the V3V4 region of the 16S rRNA gene. In addition, only DNA-based sequencing was performed by targeting both V1V3 and V3V4 regions of the 16S rRNA gene. All sequence runs included no template controls (NTC) and positive controls (*Clostridium butyricum*). Data were analyzed using the QIIME2 platform as defined in chapter 2. Sequencing analysis revealed that sequences generated from NTC could be assigned to bacterial taxa irrespective of the genetic materials and target regions, suggesting that the amplicon sequencing process introduces contaminations. When comparing the impact of the target region, alpha diversity was higher ($p<0.05$) in the fecal and oral bacterial profiles generated from the V1V3 region compared to those of the V3V4 region. Taxonomic assignment of bacterial profiles generated using two hypervariable regions revealed distinct bacterial communities. For example, Actinobacteria (fecal - $0.41\pm0.09\%$; oral - $0.51\pm0.10\%$) was abundant in bacterial profiles generated from the V1V3 region, whereas Firmicutes (fecal - $0.37\pm0.11\%$; oral - $0.34\pm0.10\%$) was abundant in those of V3V4 region when comparing D1 (prior to suckle colostrum-Day 1) samples. When comparing different genetic materials, DNA-based bacterial profiles (both oral and fecal) had a diverse microbial community compared to RNA-based profiles on D1. In contrast, the diversity of the RNA-based profiles was higher than DNA-based profiles on D2 (after suckling colostrum from cows). In conclusion, the diversity and composition of microbial communities derived from low microbial biomass samples depend on the choice of genetic materials and the hypervariable region of the 16S rRNA gene. The inclusion of appropriate controls is crucial to increase the accuracy of results, regardless of the sequencing technique.

Preface

This thesis is an original work by Ranga Nakandalage Don with the collaborations led by Dr. Leluo Guan at the University of Alberta and Dr. Nilusha Malmuthuge at Agriculture and Agri-Food Canada. The research project received research ethics approval from the University of Alberta Research Ethics Board (AUP-00004183) and the University of Saskatchewan Research Ethics Board (AUP-20170015).

Part of chapter 1 (Importance of early life microbiota) of this thesis is a part of a paper published: Nakandalage R, Guan LL, Malmuthuge N. Microbial Interventions to Improve Neonatal Gut Health. *Microorganisms*.2023;11(5): 1328.https://doi.org/10.3390/microorganisms11051328. I was responsible for the writing and original draft preparation. Guan LL, assisted with the conceptualization, writing, review, editing and funding acquisition. Malmuthuge N. was the corresponding author and was involved with conceptualization, writing, review, editing and funding acquisition.

Dedication

I dedicated this thesis to my loving parents and wife.

Acknowledgements

I would like to give my heartiest gratitude to my supervisors Dr. Nilusha Malmuthuge and Dr. Leluo Guan for their excellent guidance and mentoring throughout my masters. Their enormous guidance helps me to improve my knowledge, scientific writing, critical thinking and complete my research successfully.

I would also like to provide my sincere gratitude to Dr. Philip J. Griebel for providing me with sampling materials and for all his guidance. I am really grateful for his support and guidance to get a hand on experience in animal handling and sample collection.

I would like to thank my supervisory committee member, Dr. Anne Laarman for his support all this time and arm's length examiner Dr. Malinda Thilakarathna.

Thank you very much Yanhong Chen for all your support to conduct laboratory experiments and your hard work in DNA and RNA extraction and library preparation.

I am also thankful to all lab members for their time and support; Sang Weon Na, Anna Widenmann, Dr. Mi Zhou, Yajing Ban, Anusha Bulumulla, Yangyi Hao, Yixin Wang and Zhe Pan.

Finally, I would like to thank my loving parents and wife for their strong support and help for make my dreams come through.

Table of Contents

List of tables

List of figures

Chapter 1. Literature review

1.1 Introduction

Early life microbial colonization of the gastrointestinal tract (GIT) in human and livestock species has gained more attention due to its impact on immune modulation and overall host health (Kogut et al., 2016; Pluske et al., 2018; Caballero-Flores et al., 2022). Microbial perturbations and microbial-linked pathologies encourage the development of microbial interventional tools (Raman et al., 2019; Kim et al., 2021; Rosa et al., 2021; Song et al., 2021; Slanzon et al., 2022; Nakandalage et al., 2023; Nuzhat et al., 2023). In depth understanding of the true microbial community colonization in the GIT will facilitate the development of effective early-life microbial interventions.

Next-generation sequencing technologies (NGS) allow researchers to profile unculturable microbial communities and enhance the knowledge of microbial communities in different ecosystems (Callahan et al., 2017). However, contaminations generated during sample collection, extraction of genetic materials, PCR amplification, sequencing run, and data analysis negatively affect the identification of true microbial community when using low microbial biomass samples (Kennedy et al., 2023). Low microbial biomass samples can be defined as biological samples containing a limited amount of microbial materials to profile microbial communities (Kennedy et al., 2023). There is a higher chance to contaminate samples, if there is a low microbial material. For example, previous studies reported that the identification of microbial colonization in the fetal environment was due to the contaminations introduced from sample collection to data analysis (Eisenhofer et al., 2018; Kennedy et al., 2021; Kennedy et al., 2023).

In addition to the contaminations in low microbial biomass samples, the choice of the hypervariable region of the 16S rRNA gene affects microbial composition. Previous studies reported that microbial community composition and diversity depend on the hypervariable region of the 16S rRNA gene (Chakravorty et al., 2007; Kameoka et al., 2021; Abellan-Schneyder et al., 2021). For example, the alpha diversity of the human gut microbial community was higher in the V1V2 region compared to the V3V4 region (Kameoka et al., 2021). Chakravorty and colleagues (2007) reported that taxonomic classification differs among different hypervariable regions of the 16S rRNA gene. The genetic material used in amplicon sequencing is another important factor in microbial profiling. DNA-based amplicon sequencing profiles dead, dormant, and active microbial communities (De Vrieze et al., 2018; Salgar-Chaparro and Machuca,

2019; Wang et al., 2019). In contrast, active microbial communities can be identified using RNA-based amplicon sequencing (De Vrieze et al., 2018). Thus, the hypervariable region of the 16S rRNA gene and genetic materials for microbial profiling have to be selected cautiously. However, there is limited knowledge on suitable hypervariable regions of the 16S rRNA gene and genetic material to identify fecal and oral microbial community in newborn beef calves, when using swabs to collect samples.

The introduction of NGS technologies has not only increased the knowledge of microbial communities but also encouraged researchers to develop microbial analysis platforms to interpret microbial profiling outcomes (Callahan et al., 2016). However, the data analysis pipeline itself affects the outcomes of microbial community assessment based on amplicon sequencing. Thus, the credibility of microbial data interpretation depends on the proper use of different bioinformatics tools. However, the customization of bioinformatics pipelines to obtain credible outcomes in microbial profiling is limited.

1.2 Importance of early life microbiota[1]

1.2.1   Early life microbiota and gut health

Mammalian gut health has gained research and public attention over the last decade due to its influence on overall host health (Kogut et al., 2016; Pluske et al., 2018). Traditionally, gut health has been defined as the ability to maintain disease- or pathogen-free status (Bischoff, 2011). However, gut health includes several physiological and functional features of the gastrointestinal tract (GIT) including nutrient digestion and absorption, host metabolism, microbiota, barrier function, and mucosal immune responses (Bischoff, 2011; Kogut et al., 2016; Pluske et al., 2018). Therefore, gut health can be defined as a multifactorial concept that depends on the host and microbial interactions to maintain metabolism, homeostasis, immune functions, and overall wellbeing.

Host–microbe interactions are one of the primary elements of gut health (Malmuthuge and Guan, 2017; Ravisankar et al., 2018; Swanson et al., 2020), which can be affected by various internal and external factors throughout life. Birth mode (vaginal delivery vs. cesarean delivery), gestational age (premature vs. full-term), antibiotic treatments, diet (breastfed vs. formula-fed; fiber-rich vs. high-sugar, high-fat diet),

---

[1] Importance of early life microbiota is a part of a paper published: Nakandalage R, Guan LL, Malmuthuge N. Microbial Interventions to Improve Neonatal Gut Health. *Microorganisms*.2023;11(5):1328.https://doi.org/10.3390/microorganisms11051328.

infections, habits (exercise, stress), and environmental factors (geographical region) have been linked to perturbed microbiota and microbial-linked pathologies (Yatsunenko et al., 2012; Danzeisen et al., 2015; Malmuthuge et al., 2015; Ballou et al., 2016; Bamuler and Sperandio et al., 2016; Vogt and Finlay, 2017; Litvak et al., 2018; Ho et al., 2018; Li et al., 2021). However, some of these factors that lead to microbial dysbiosis are life-saving medical interventions and are thus indispensable. For example, despite the vast literature supporting the presence of microbial perturbations in cesarean-delivered babies and the increased risk of developing microbiome-linked pathologies later in life, it has tremendously improved the survival of babies and moms by reducing the number of deaths due to complications during pregnancy (Geleto et al., 2020). In such situations, microbial interventions can be used to restore perturbed microbial communities and to mitigate negative health outcomes/microbiome-linked pathologies.

The colonization of gut microbiota starts during the birthing process when the fetus is exposed to the outside following the rupture of amniotic membranes (Caballero-Flores et al., 2023). In addition to infants, recent studies on livestock have shown that early life events are linked to variations in the microbial community. For example, naturally born calves had a higher microbial richness, evenness, and diversity in the rumen when compared to cesarean-delivered calves (Furman et al., 2020). Moreover, the time of colostrum feeding has been shown to affect the establishment of the microbial community in newborn calves (Ma et al., 2019). In piglets, antibiotics have been reported to alter the gut microbiota community (Schokker et al., 2014; Li et al., 2018; Correa-Fiz et al., 2019; Fouhse et al., 2019; Guevarra et al., 2019). Moreover, piglets raised in isolators and treated with antibiotics were reported to have altered microbial communities and immune responses (Mulder et al., 2009).

1.2.2   Early life microbiota and immune modulation

Maintaining a balance between protective and regulatory immune responses is vital for neonatal gut health. Neonates are particularly susceptible to infections, until their immune systems are functionally developed (Zhang et al., 2017; Hornef and Torow, 2020; Sangild et al., 2021). Germ-free animal models have revealed the causal relationship between commensal gut microbiota and the development of the host immune system during early life. For example, germ-free mice have fewer B cells (Hansson et al., 2011; Maynard et al., 2012; Zitvogel and Kroemer, 2021) and lower levels of secretory antibodies and antimicrobial peptides when compared to conventional mice (Hansson et al., 2011; Lamouse-smith et al., 2011;

Gensollen et al., 2016; Thomson et al., 2022). Moreover, studies have shown that preterm infants with perturbed gut microbial communities lack goblet cells, Paneth cells (McElroy et al., 2011), and natural killer cells (Zhou et al., 2016) compared to full-term infants (Collado et al., 2015; Xiang et al., 2022). This immune memory activated by the pioneer microbiota is known as immune imprinting (Al Nabhani et al., 2020). When the pioneer microbial community is perturbed, the interactions between host and gut microbiota can lead to the overactivation of reactive immune responses, disturbing immune homeostasis (Yamamoto et al. 2012; Al Nabhani et al., 2020). However, these altered immune functions can only be returned by restoring microbiota during the neonatal period (birth to weaning) (Gensollen et al., 2016), indicating that microbial interventions to modulate the immune system should be conducted during this window of opportunity. In germ-free mice, the accumulation of invariant natural killer T cells (iNKT) in the colon and lung increased susceptibility to inflammatory bowel disease and asthma (Olszak et al., 2012). However, the colonization of germ-free mice before weaning (in the neonatal period) could only reduce iNKT numbers and disease incidence (Olszak et al., 2012). The recognition of lipid antigens (self and non-self) by iNKT cells, which establish a long-term residency in tissues, activates pro-inflammatory cytokines and skewed immune responses towards inflammation (Loh et al., 2014; Crosby et al., 2018). Inflammation is one of the key signatures of microbiome-linked pathologies, and early-life microbial colonization plays a vital role in maintaining immune homeostasis by modulating the development of the immune system. Similarly, germ-free mice have fewer mucosal-associated invariant T cells (MAIT) than specific-pathogen-free mice (Constantinides et al., 2019). When germ-free mice were colonized during early life with a cocktail of commensal bacteria from wild-type mice, it restored the number of MAITs, but this was not the case in adult mice (Constantinides et al., 2019). These studies suggest that immune imprinting occurs when introducing microbiota during the neonatal period but not after. Thus, the neonatal period is a crucial time window to intervene in the gut microbiota to maturate immune functions and ensure a healthy gut environment later in life.

1.2.3 Early life microbiota and gut epithelial barrier functions

In addition to priming balanced immune responses, gut microbiota also plays a vital role in maintaining gut epithelial barrier functions (Gonzalez-Gonzalez et al., 2018; Ghosh et al., 2021). Appropriate regulation of barrier functions is another aspect of gut health (Bischoff, 2011). High intestinal permeability is also

common in preterm infants due to their immature gut microbiota composition (Ma et al., 2022). However, *Bifidobacterium breve* and human milk oligosaccharides decrease intestinal permeability in preterm infants (Ma et al., 2022a). A recent study in piglets revealed that antibiotic treatments for diarrhea decrease gut permeability (Ma et al., 2022b). In this study, administration of ampicillin for three days decreased the relative expression of tight-junction and adherence-junction proteins in the colons of newborn piglets. The expression of these proteins increases following fecal microbial transplantation (FMT), indicating that microbial restoration improves barrier functions. In neonatal calves, feeding colostrum improves the gut barrier integrity compared to feeding formula by increasing the expression of tight-junction proteins at four days of age (Ghaffari et al., 2021). Feeding colostrum has also been shown to increase the colonization of beneficial bacteria such as Lactobacillus and Bifidobacterium in the GIT of newborn calves (Malmuthuge and Guan, 2017; Fischer et al., 2018; Song et al., 2019). However, knowledge is lacking regarding whether the beneficial changes in the gut microbial community due to colostrum feeding play a role in the improved barrier functions of calves. Neonatal calf diarrhea is one of the major concerns in dairy calves, as it increases gut permeability (Araujo et al., 2015) and alters gut microbial community (Kim et al., 2021). Recently, FMT has been used to minimize neonatal calf diarrhea and to improve gut health in pre-weaned calves (Kim et al., 2021; Rosa et al., 2021; Slanzon et al., 2022). These studies reported that FMT successfully altered gut microbial community. However, the impact of FMT on intestinal barrier functions and gut permeability is yet to be understood. The use of probiotic supplements in dairy calves has been shown to increase the expression of tight-junction gene zonula occludens-1 and occludin, while increasing microbial diversity (Wu et al., 2021). Therefore, studying the impact of FMT on gut barrier functions will explain the modulatory mechanisms behind reduced diarrhea in dairy calves during FMT. However, the use of microbial intervention in newborn beef calves have not been studied due to the limited understanding of microbial colonization and difficulties in sample collection from newborn beef calves. These studies highlight the importance of beneficial early-life microbiota in regulating intestinal permeability. A well-maintained gut barrier is crucial for neonates to maintain gut homeostasis. Thus, it is evident that early-life microbiota plays a vital role in maintaining gut health by modulating immune functions, cellular populations, and barrier integrity.

1.3 Profiling microbial community using low microbial biomass samples

Low microbial biomass samples can be defined as biological samples containing a limited amount of microbial materials to profile microbial communities, such as oral, nasopharyngeal, and skin swabs, tissue and mucosal samples, and bodily fluids (Kennedy et al., 2023). Swabbing is one of the non-invasive sampling techniques that can perform, transport, and store with less complexity (Reyman et al., 2019; Radhakrishnan et al., 2023). Moreover, it allows repetitive sampling of the same animal to study the dynamics of microbial communities over a period. A study comparing human gut microbial composition between rectal swabs and fecal samples reported a higher correlation between the two sampling methods (Radhakrishnan et al., 2023), suggesting the potential of using swabs for sample collection. The collection of fecal samples can be challenging in newborns due to the availability of fecal samples immediately after birth (Reyman et al., 2019). Thus, swab samples are one of the ways to profile the microbial community. However, there are limitations when profiling low microbial biomass samples (swab samples) due to the low amount of microbial materials.

1.3.1   Importance of studying microbial community using low microbial biomass samples

Rapid advancements in next-generation sequencing (NGS) have improved the knowledge of microbial community composition. However, the accuracy of the data generated through NGS and user-friendly bioinformatics tools has generated debatable outcomes. For example, the presence of microbiota in the fetus and intrauterine environment during a healthy pregnancy is highly debatable in microbiome research (Lauder et al., 2016; Lim et al., 2018; Malmuthuge and Griebel, 2018; De Goffau et al., 2019; Guzman et al., 2020; Rackaityte et al., 2020; Stinson et al., 2020; Kennedy et al., 2021; Mishra et al., 2021). Previous studies conducted to identify fetal microbial community in C-section and vaginal-delivered fetuses (terminated pregnancies) identified a low number of microbial taxa by sequencing the 16S rRNA gene (Guzman et al., 2020; Rackaityte et al., 2020; Stinson et al., 2020; Mishra et al., 2021). For example, one of the studies reported that 18 bacterial taxa were enriched in the vaginally-delivered terminated fetuses during the second trimester compared to negative controls (Rackaityte et al., 2020). Sequencing of the 16S rRNA gene to profile microbiota across fetal organs detected the low number of microbial taxa in the gut, skin, placenta, and lungs after elective termination of pregnancy during the second trimester (Mishra et al., 2021). In contrast, sequencing of fetal meconium samples collected by swabbing during cesarean surgeries

before delivery confirmed contamination taxa in negative controls (Kennedy et al., 2021). De Goffau and colleagues (2021) reanalyzed the fetal data from a study conducted by Rackaityte and colleagues (2020) and reported that *Micrococcus luteus* identified in fetal samples were falsely assigned to fetal samples due to the batch effect. Rackaityte and colleagues (2020) used two different batches of samples by changing the sampling procedure and including negative control only in the second batch. During bioinformatics data analysis using decontam to remove contaminations cannot distinguish *Micrococcus* from fetal samples in batch 1. However, *Micrococcus* was not present in batch 2 fetal samples after removal of contaminations using decontam due to the inclusion of negative controls in batch 2, suggesting the importance of including controls in low microbial biomass studies. A recent study conducted by Kennedy and colleagues (2023) clearly explained that the microbial signals detected in fetus samples resulted from contaminations during sampling, nucleic acid extraction, and sequencing. In addition to humans, Malmuthuge and Griebel (2018) confirmed the *in-utero* sterility of the ovine fetus during the third trimester of pregnancy when using appropriate controls and proper sample collection methods. All these studies suggest that contaminations and cross-contaminations are challenges of sequence-based microbiome research, especially when using samples with low or no microbial biomass. Early-life microbial colonization affects gut health, barrier functions, and immune imprinting (Nakandalage et al., 2023). The generation of accurate knowledge on the early life microbial colonization process is important for the development of microbial interventional tools to overcome the lifelong consequences of microbial-linked pathologies and maintain overall host health.

1.3.2 Pitfalls in profiling microbial community using low microbial biomass samples

When using low microbial biomass samples, contaminations originating during microbial profiling outweigh the true microbial signals compared to high microbial biomass samples (Kennedy et al., 2023). Contaminations can occur from the beginning of sample collection until data analysis. Previous studies reported that contaminations originated through samples, laboratory environments, laboratory reagents including PCR master mixes, nucleic acid extraction kits, plastic consumables, and researchers (Table 1.1). For example, a study conducted to identify the contaminations in commonly used DNA extraction kits reported that microbial signals could be detected in all DNA extraction kits used in the study and the microbial profiles varied between different kits and batches (Salter et al., 2014). Another study identified 88 bacterial genera when using molecular-grade water to extract DNA (Glassing et al., 2016). Moreover, cross-

contaminations can occur through tag switching (barcode cross-contamination), index hopping (mismatch indexing reads to sequencing reads), residual sequences from previous sequencing runs, and genetic materials from other samples during sample processing (Carlsen et al., 2012; Eisenhofer et al., 2018; Larsson et al., 2018). These studies suggest that microbial communities are misrepresented by these potential contaminations, especially when microbial DNA is low in the samples. The high sensitivity of NGS technologies helps to detect contamination and/or cross contaminations, if include proper controls in microbial profiling.

Furthermore, artificial sequences generated through sequencing runs and data analysis pipelines lead to misinterpretation of microbial community compositions. Different algorithms and plugins in bioinformatics platforms have been used to remove contaminations from microbial data. For example, DADA2, deblur, cutadapt, trimmomatic are a few different bioinformatics tools used to remove sequencing errors, primers, and barcode sequences using different algorithms (Chen et al., 2014; Callahan et al., 2016). However, the use of default settings in computational quality control tools (e.g., denoising, quality filtering) leads to false positive interpretations in low or no-microbial biomass research (Davis et al., 2018), suggesting the importance of optimization of bioinformatics tools to obtain credible outcomes. A study conducted by Malmuthuge and Griebel (2018) used a length-based quality filtering step to remove non-specific sequences from the 16S amplicon sequence data when the use of gel-based amplicon purification was not available. Authors identified cross contaminations in positive controls and contaminated taxa in the fetal intestine in the absence of the length-based quality filtering step (Malmuthuge and Griebel, 2018). This study suggests that the data analysis pipeline itself can introduce contaminations during microbial community profiling. Rai and colleagues (2021) reported that different truncation lengths during denoising affect the downstream analysis. Their findings revealed that removing low-quality sequences influences the sample size, which affects the statistical power during downstream analysis (Rai et al., 2021). Another study compared microbial diversity and composition with and without a denoising step (denoising removed noises identified during pyrosequencing and qPCR) reported reduced accuracy of alpha diversity analysis in the absence of denoising (Reeder and Knight, 2010). Thus, appropriate data analysis pipelines are required to maintain the credibility of data obtained from low microbial biomass samples. This created a

need to optimize bioinformatics tools based on the quality of the data set to identify real/true microbial communities beyond the use of appropriate controls to identify contaminations.

1.3.3 Mitigating the challenges in profiling microbial community using low microbial biomass samples

Contaminations can be controlled throughout the microbial profiling process including; 1) during sampling, 2) amplicon generation, 3) sequencing, and 4) data analysis.

1.3.3.1 Preventing contaminations during sampling

Careful sample collection and processing is one of the ways to overcome contaminations and biases in low microbial biomass samples. Sample collection can introduce contaminations from the environment, laboratory surfaces, and humans. These contaminations can be minimized by performing proper sample collection and handling. For example, a study conducted to profile microbiota in ovine fetuses during the third trimester of pregnancy maintained sterility during sampling (Malmuthuge and Griebel, 2018). In this study, researchers maintained a sterile surgical field, autoclaved surgical instruments, immediate sample collections after opening fetal membranes, avoid contaminations from the skin, and minimized human microbiota contamination by minimizing the number of people involved in the surgical procedure, wearing sterile masks, gloves, hair covers and gowns (Malmuthuge and Griebel, 2018). These strict guidelines during sample collection and handling helped the authors confirm fetal gut sterility during the third trimester of pregnancy (Malmuthuge and Griebel, 2018). In addition, Eisonhofer and colleagues (2019) suggested the importance of using the same personnel to collect the samples, the same reagents and equipment to perform laboratory experiments, wearing protective clothes and equipment to minimize human contamination, and use of reagents and consumables with the lowest level of contaminations during sampling (Eisenhofer et al., 2019).

1.3.3.2 Preventing contaminations during amplicon preparation

The use of proper controls during amplicon preparation to identify contaminations and cross-contaminations can reduce the generation of controversial findings in microbial research. Previous studies that identified microbial communities in no/low microbial biomass samples did not report the use of appropriate controls (Satokari et al., 2009; Aagaard et al., 2014; Doyle et al., 2014; Collado et al., 2016; Nagpal et al., 2016; Zhu et al., 2018). However, the use of negative controls and blanks during the sequencing of no or low

microbial biomass samples facilitated the identification of contaminated microbial taxa from commercial DNA extraction kits, molecular grade reagents, and the environment (Kennedy et al., 2023). Lauder and colleagues (2016) reported that the microbial profiles of placental samples cannot be distinguished from those of blank controls. They found that contaminations introduced during DNA purification could not be separated from placental samples (Lauder et al., 2016). Glassing and colleagues (2016) reported 181 contaminated bacterial genera in molecular-grade reagents. These studies clearly suggest the importance of incorporating appropriate controls in all steps of microbial profiling. As a result, recent studies used various types of controls to monitor the contaminations and cross-contaminations, such as sample blanks, DNA extraction blanks, no template controls during amplification, and a known bacterial community as a positive control (Eisenhofer et al., 2019). Negative controls help to monitor background contamination levels and understand the sources of contaminations. This helps remove contaminations before downstream data analysis. For instance, the use of negative controls in studies with low microbial biomass samples detected 60 microbial taxa across multiple studies (Eisenhofer et al., 2019). The detected microbial taxa in negative controls may depend on the nucleic acid extraction kit, molecular grade reagents, and sequencing methods. Cross-contamination of biological samples can be detected by including positive controls. Researchers use mock communities or serial dilution of known bacterial taxa as positive controls to identify cross-contaminations (Minich et al., 2018; Eisenhofer et al., 2019). These controls must run alongside biological samples to detect contaminations and cross-contaminations.

1.3.3.3 Preventing contaminations during sequencing

During sequencing, hundreds of libraries are sequenced parallelly on one lane of a flow cell, which can lead to cross-library contaminations during high-throughput next-generation sequencing (Larsson et al., 2018). Cross-contamination from previous sequencing runs is another way to introduce contamination during sequencing (Eisenhofer et al., 2018). Maintenance wash in between sequencing runs should be performed to avoid run-to-run cross contaminations in sequencing machines (Eisenhofer et al., 2018). Barcode sequencing errors and index hopping (mismatch of index reads to sequencing reads) are other ways of cross-contamination on the sequencing instrument (Larsson et al., 2018). The use of appropriate controls (negative and positive controls) that went through the amplicon generation step at the same time as

samples allow for identifying cross contaminations and contaminations from sequencing reagents and flow cells (Eisenhofer et al., 2018).

1.3.3.4 Preventing contaminations during data analysis

Data analysis pipelines and bioinformatics platforms also influence the outcomes of microbial profiling. Comparison of biological samples with controls helps to determine the contaminated microbial taxa (Eisenhofer et al., 2019). Saffarian and colleagues (2019) reported that removal of contamination taxa identified in previous studies during bioinformatics analysis without using proper controls to identify contaminations in the present study might lead to misinterpretation of microbial communities. These finding suggest the importance of using appropriate controls to identify contaminations in each study. The denoising step in the QIIME2 bioinformatics pipeline helps to remove low-quality raw sequences (barcodes, primer dimers, chimeras, PCR errors and sequencing errors) using DADA2 or deblur denoising algorithms (Hall and Beiko, 2018). However, the denoising step cannot be used to remove contaminated taxa. Therefore, quality filtering steps should be customized based on the quality of the data set after sequencing. For example, ASVs/OTUs identified in no template controls can be filtered out by quality filtering following denoising to remove contaminations from reagents.

In addition, Decontam is an open-source R package used to identify contaminations in amplicon and metagenomics sequence data by using a statistical approach (Davis et al., 2018). Decontam identifies and removes contaminated ASVs/OTUs/taxa present in negative controls based on the relative abundance and frequency. Decontam use two types of metadata (1) DNA concentration and (2) sequenced negative control samples to remove contaminations (Davis et al., 2018). These data allow decontam to perform statistical analyses and identify contaminations based on the prevalence and frequency of microbial samples in negative controls. However, decontam neither recognizes the presence of contamination when contaminations present in few samples nor removes cross-contaminations (Davis et al., 2018). Decontam identify contaminations based on pattern of samples. Therefore, decontam is less sensitive to detecting contaminations presence in low number of samples (Davis et al., 2018), suggesting that decontam may not applicable for low microbial biomass samples.

1.4 Molecular techniques to profile microbiota

1.4.1 Methods to profile microbial communities

Currently, there are a number of tools available to profile microbiota. Culturing of bacteria and microscopic evaluations were the first methods employed to report microbes in an ecosystem (Lagier et al., 2015). These microbial detection methods have become the gold standard for detecting microbiota for many years (Fraher et al., 2012). Biochemical characterization of novel microbes can fully understand if they can grow in a culture media and study *in vitro* (Fraher et al., 2012). This is one of the main advantages of culture-based methods, which is important to understand the characteristics of novel microbes. To date, culture-based approaches characterized less than 30% of microbiota (Fraher et al., 2012). The limited understanding of the growth conditions of anaerobic microbes negatively affects the development of proper culture media to grow them *in vitro* (Fraher et al., 2012). Thus, culture-based approaches can generate only a limited knowledge of the GIT microbiota, which is dominated by anaerobes. Culture-based microbial identification can also be time-consuming.

Genetic materials of microbes have been used to profile microbial communities in culture-independent molecular-based microbial profiling approaches. The use of genetic materials has provided the opportunity for the rapid identification of both culturable and non-culturable organisms. There are a number of ways to profile microbial communities using culture-independent approaches (Table 1.2). For example, denaturing gradient gel electrophoresis (DGGE), terminal restriction fragment length polymorphism (T-RFLP), fluorescence *in situ* hybridization (FISH), and DNA microarrays were used as low-throughput microbial profiling techniques in the past (Fraher et al., 2012; Strathdee and Free, 2013). DGGE has provided valuable information on microbial communities by separating short to medium-length DNA fragments (Fraher et al., 2012; Strathdee and Free, 2013). However, the use of the DGGE technique in microbial profiling is limited due to the less reproducibility, labor intensiveness, and the availability of limited sequence information (Fraher et al., 2012; Strathdee and Free, 2013). T-RFLP has the same limitation when profiling the microbial community, even if it provides a quantitative comparison through visualization of DNA fragments on a gel by fluorescent labeling (Fraher et al., 2012). Probe hybridization techniques, such as FISH and DNA microarrays use fluorescent-labeled oligonucleotide probes to target bacterial DNA sequences, which allows phylogenetic identification of species (Fraher et al., 2012). However, these

techniques cannot be used to target unknown bacterial species (probes designed to target specific bacterial species) and there is limited detection of low-abundance species. Even though these techniques improved the knowledge of microbiota colonization in different ecosystems, limitations including no phylogenetic identification, inability to identify unknown species, and difficulties in identifying low abundance species limit the use of these techniques for the characterization of microbial communities.

The first-generation sequencing approach, Sanger sequencing has been widely used until the 90th century to identify cultured and uncultured bacteria with phylogenetic relationships using chain termination by dideoxynucleotides (Jamuar et al., 2016; Mardis, 2017). However, due to the high cost and labor-intensive process of Sanger sequencing, next-generation sequencing (NGS) has been introduced to reduce the cost and enhance the output of sequencing. NGS techniques generate high-throughput data from a large number of samples in parallel within a short period. Illumina, 454 Pyrosequencing, and SOLiD are commercially available next-generation sequencing technologies (Rothberg and Leamon, 2008; Ellermann et al., 2017). Among NGS techniques, 454 Pyrosequencing was the first available NGS technique to profile microbiota (Rothberg and Leamon, 2008; Bleidorn, 2017). It can sequence a higher number of sequences (500 million bases) in a single run compared to Sanger sequencing (Fraher et al., 2012). However, the high cost and high error rates are limitations of 454 pyrosequencing (Fraher et al., 2012; Razali et al., 2017). Illumina currently dominates microbial sequencing research due to the higher read depth and coverage compared to other sequencing techniques. Illumina sequencing allows the generation of high throughput read counts with lower cost and error rates (Luo et al., 2012; Bleidorn, 2017).

Amplicon and metagenomics sequencing are used widely to profile microbial communities depending on the research objectives. Amplicon sequencing uses marker genes to profile microbiota. The 16S rRNA gene is mainly used to profile bacteria and archaea communities, while internal transcribed spacer (ITS) is used to profile fungal communities (Nilsson et al., 2009; Johnson et al., 2019). The presence of the 16S rRNA marker gene in all bacteria species provides an opportunity to use it as a universal marker gene (Wang et al., 2015). In contrast, metagenomics sequencing of total DNA extracted from a microbial community generates knowledge of both taxonomic and functional compositions (Munshi and Sharma, 2018). The functionality of a microbial community is explained as the abundance of genes and functions (Gilliland et al., 2012; Fraher et al., 2012). Metagenomics sequencing can be costly and contaminated by

host and environmental DNA. Moreover, metagenomics sequencing requires a higher amount of DNA as the starting material (Quince et al., 2017; Knight et al., 2018). Thus, metagenomics shotgun sequencing is not applicable for samples that are low in microbial biomass or degraded (Quince et al., 2017; Knight et al., 2018). Metatranscriptomics sequencing can also be used to assess the functionality of a microbial community. This approach uses RNA extracted from a microbial community and assesses the expression of microbial genes (Jovel et al., 2022).

1.4.2 Targeting hypervariable regions of the 16S rRNA marker gene to profile microbial communities

The 16S rRNA marker gene is around 1600 bp long and phylogenetically conserved across all prokaryotic species (Bharti and Grimm, 2021). Thus, the 16S rRNA marker gene has been considered the gold standard marker gene to study microbial ecology. It consists of nine hypervariable regions (V1-V9) with conserved regions in between the variable regions, which can be used to design primers to target phylogenetically different taxa (Chakravorty et al., 2007; Abellan-Schneyder et al., 2021). However, all hypervariable regions have not displayed the same sensitivity and resolution in terms of profiling microbiota (Yang et al., 2016). Thus, no single hypervariable region can be used to distinguish among all bacterial species. Besides, one hypervariable region may not be suitable to profile microbial communities from various ecosystems due to the distinct microbial communities among different ecosystems

Taxonomic classification varies when targeting different hypervariable regions using different primer sets. One of the studies conducted by Chakravorty and colleagues (2007) characterized 110 pathogenic bacterial species using primer sets to target V1-V8 hypervariable regions. They revealed that the V1 region of the 16S rRNA gene differentiated *Staphylococcus* spp. However, V2 and V3 regions were suitable to distinguish all pathogenic bacterial species to genus level except, species belonging to *Enterobacteriaceae* (Chakravorty et al., 2007). In addition, the taxonomic resolutions were lower in V4, V5, V7, and V8 hypervariable regions up to genus or species level (Chakravorty et al., 2007). This suggests that primers targeting different hypervariable regions have varying capacities to profile bacterial communities. Moreover, primer sets covering more than one variable region in the 16S rRNA marker gene can improve the resolution of microbiota profiling (Abellan-Schneyder et al., 2021). The fecal bacterial community composition of the same individual was distinct when sequenced using different primer sets. For instance, the use of seven

different primer pairs targeting seven hypervariable regions (V1-V2, V1-V3, V3-V4, V4, V4-V5, V6-V8, and V7-V9) of the 16S rRNA gene to profile human fecal microbiota revealed that *Verrucomicrobia* was only detected when using V3-V4, V4, V6-V8, and V7-V9 primers, while *Proteobacteria* and *Firmicutes* phyla can be identified from all primer pairs (Abellan-Schneyder et al., 2021). The presence and absence of phyla *Actinobacteria, Tenericutes, and Lentisphaerae* differed across hypervariable regions (Abellan-Schneyder et al., 2021), confirming that all bacterial phyla cannot be detected using a single hypervariable region of the 16S rRNA gene. Moreover, the V4-V5 primer pair displayed a low abundance of *Bacteroides* (Abellan-Schneyder et al., 2021). This dissimilarity between different hypervariable regions of the 16S rRNA gene is due to the different resolutions of different hypervariable regions to profile bacterial taxa. For example, some bacterial taxa are unclassified at the genus level when using some of the 16S rRNA hypervariable regions, whereas the same bacterial taxa can be classified when using different hypervariable regions of the 16S rRNA gene. This is one of the reasons for the observed higher variability of microbial taxa between different regions of the 16S rRNA gene (Abellan-Schneyder et al., 2021). *In Silico, an* assessment of the theoretical coverage of bacterial genera using the Silva database revealed a higher coverage in V1-V2, V1-V3, and V3-V4 compared to V6-V8 and V7-V9 regions (Abellan-Schneyder et al., 2021). These results suggest that the choice of primer pairs targeting the hypervariable region of the 16S rRNA gene is one of the important factors to consider in microbiota profiling.

The size of the amplicon (amplicon length) also influences the resolution of taxonomic classification. A higher amplicon length in microbial profiling leads to generating higher taxonomic resolution of microbial community composition. For example, a study comparing the microbial profiles of the human gut generated using full-length (V1-V9) and short-read (V3-V4) amplicons reported a higher resolution at the species level when using full-length amplicon sequencing (Matsuo et al., 2021). Second-generation sequencing (Illumina sequencing) provides the opportunity to sequence amplicon lengths up to 600 bp using Illumina MiSeq sequencing technology. However, third-generation sequencing enables to sequence of full-length 16S rRNA genes using Oxford Nanopore MinION and PacBIOs Sequel (Bharti and Grimm, 2021). This allows the sequence up to 10,000 bp long-length reads in a short time (Bharti and Grimm, 2021). However, relatively higher error rates, high costs, limited applicability for high throughput techniques, and less standardization of protocols and data analysis pipelines are major drawbacks in third-generation

sequencing (Bharti and Grimm, 2021). On the other hand, 16S rRNA gene amplicon sequencing is more compatible with samples contaminated with host (tissue samples) and low microbial biomass (Knight et al., 2018).

1.4.3 Choice of genetic materials to profile microbial communities

Microbial community composition depends on the type of genetic material used to generate amplicons. Amplicon sequencing-based microbial profiling mainly relies on DNA as the genetic material to profile microbiota in an ecosystem. However, DNA-based diversity profiling identifies the total microbial community of a niche, including dead, active, and dormant microbial cells (De Vrieze et al., 2018; Salgar-Chaparro and Machuca, 2019; Wang et al., 2019). Dead and dormant microbial cells lead to an overestimation of active microbial communities (Wang et al., 2019). Previous studies reported that not all microbiota in a community are active and the active microbes cannot be differentiated through DNA-based approaches (Salgar-Chaparro and Machuca, 2019). In contrast, RNA-based amplicon sequencing can only profile active microbes in a community (De Vrieze et al., 2018). Previous studies comparing DNA and RNA-based bacterial density and diversity revealed significant differences between total and active microbial communities (Li et al., 2016; De Vrieze et al., 2018; Malmuthuge et al., 2019; Salgar-Chaparro and Machuca, 2019). One of the studies comparing the microbiota in oil production facilities revealed significant differences in the microbial community between DNA and RNA-based diversity profiling (Salgar-Chaparro and Machuca, 2019). Authors reported that the richness of the microbial community was higher in RNA-based profiling compared to DNA-based profiling (Salgar-Chaparro and Machuca, 2019). A previous study conducted by Li and colleagues (2016), compared the rumen microbial community composition of beef steers using DNA and RNA-amplicon sequencing and revealed a higher alpha diversity in RNA-based amplicon sequencing. Another study conducted by Malmuthuge and colleagues (2019) quantified rumen bacterial densities from birth to six weeks of life using DNA and RNA and reported higher rumen bacterial densities in DNA-based approaches compared to RNA-based approaches from day zero to six weeks in dairy calves (Malmuthuge et al., 2019). Another study reported that archaea communities of anaerobic digestion plants showed a significantly higher alpha diversity on DNA-based sequencing compared with RNA-based sequencing (De Vrieze et al., 2018). Similarly, beta diversity analysis based on a weighted UniFrac distance matrix reported that archaea communities between DNA and RNA-based sequencing

were significantly different (De Vrieze et al., 2018), suggesting the difference in archaea community composition. However, they did not observe significant differences in diversity for the bacterial community between DNA and RNA-based sequencing. Moreover, this study (De Vrieze et al., 2018) did not mention information regarding the PERMANOVA analysis (e.g., R^2) to show how much genetic material contributed to the separation observed in beta diversity analysis. In addition, the relative abundance of some rumen microbial families (e.g., *Streptococcaceae*) was higher in DNA-based amplicon sequencing, while the relative abundance of some bacterial families (e.g., *Desulfovibrionaceae, Elusimicrobiaceae* and *Sphaerochaetaceae*) were higher in RNA-based amplicon sequencing (Li et al., 2016), suggesting that changes in microbial community composition among DNA and RNA-based microbial profiling impact on microbial diversity. In addition, some bacterial families have not been detected in DNA-based sequencing (Li et al., 2016). For example, two bacterial phyla (*Elusimicrobia* and *Verrucomicrobia*) and one bacterial family (*Elusimicrobiaceae*) were only identified in RNA-based sequencing (Li et al., 2016). The difference between DNA and RNA-based microbial profiling in terms of diversity might be due to the identification of dead microbial cells in DNA-based profiling, which masks the lower abundant active microbial cells. This difference highlights the importance of using RNA-based approaches to identify active microbial communities in an ecosystem. Even though conducting the RNA extraction procedures is more laborious due to the susceptibility to degradation, RNA-based approaches generate a better understanding of active microbiota (Salgar-Chaparro and Machuca, 2019). In addition to the identification of dead and dormant microbial cells by DNA-based profiling, microbial diversity, and community composition are overestimated by DNA-based approaches due to the presence of more than one 16S rRNA gene copy in some bacterial genomes (Kembel et al., 2012; Angly et al., 2014; Louca et al., 2018). However, the difference between total (DNA-based) and active (RNA-based) microbial communities in newborn beef calves has not been well studied previously using low microbial biomass samples.

1.4.4   Microbial profiling and bioinformatics tools

The introduction of NGS technologies not only lead to studying various microbial communities but also encouraged researchers to develop novel and user-friendly microbial analysis platforms to summarize and interpret data (Callahan et al., 2016). High throughput data should be handled properly to generate credible outcomes from microbial studies. There are many open-source software for microbiome data analysis,

including Mothur, Quantitative Insights into Microbial Ecology (QIIME2; https://QIIME2.org), and phyloseq (Schloss et al., 2009; Reeder and Knight, 2010; Golob et al., 2017; McMurdie and Holmes, 2013; Karstens et al., 2018; Bolyen et al., 2019). These bioinformatics tools convert millions of sequencing reads into operational taxonomic units (OTUs) or ASVs, followed by taxonomic classifications (Karstens et al., 2018). A combination of bioinformatics tools creates a bioinformatics pipeline to convert raw sequences into microbial profiles. However, a default workflow of developers might not fit all microbial data analysis.

Bioinformatics tool is one of the factors that determine the outcomes of the microbiota surveys. For example, one of the recent studies used different bioinformatics pipelines to assess how each bioinformatics pipeline can affect biological conclusions (Siegwald et al., 2019). Interestingly, authors identified significant differences in microbial community richness and diversity when comparing Mothur and Qiime bioinformatics pipelines (Siegwald et al., 2019). For example, lower richness and diversity were observed when using mothur compared to QIIME2 (Siegwald et al., 2019). Another study reported that bioinformatics platforms and reference databases affect diversity analysis of microbial communities (Golob et al., 2017). They found that *in silico* generated bacterial community analysis using the QIIME2 bioinformatics platform correctly classified 12% of organisms when using greengenes database, while the Silva database correctly classified only 8.8% of organisms (Golob et al., 2017). Marizzoni and colleagues (2020) compared, four commonly used user-friendly bioinformatics pipelines (QIIME2, Bioconductor, UPARSE, and mothur) using human fecal samples to evaluate the impact of bioinformatics pipelines on the taxonomic classification. Authors reported that taxonomic assignment at phylum and genus level was similar across all the pipelines while showing differences in terms of relative abundance for all phyla identified (Marizzoni et al., 2020). Among these microbial data analysis software, QIIME2 provides an interactive platform to conduct data analysis, visualization, and statistical analysis through one open-source software package (Mohsen et al., 2019). Mothur can only be used to perform data analysis but users have to rely on other software (eg: R statistical software package) to visualize data (Bolyen et al., 2019). Phyloseq tools conduct statistical analysis of microbial data and generate visualizations (Bolyen et al., 2019). Thus, Phyloseq can be used to analyze outputs generated from QIIME2 and Mothur.

Due to its user-friendly nature (Caporaso et al., 2010; Bolyen et al., 2019), QIIME2 has been widely used for microbiota analysis in a wide range of research areas. It contains a series of plugins (Figure 1.1),

including demultiplex (demux), denoising (dada2/deblur), generation of the feature table, and taxonomic classification (feature-classifier), and diversity analysis (diversity) to convert raw sequencing data into graphics-based results (Mohsen et al., 2019). For example, the "q2-demux" plugin in QIIME2 allows the creation of an interactive sequence quality plot to visualize the quality of sequencing data (Hall and Beiko, 2018). The interactive sequence quality plot is a box plot that visualizes the quality of each base position (Hall and Beiko, 2018). These interactive quality plots are used in determining the truncation length of sequencing reads in subsequent denoising analysis (Hall and Beiko, 2018). However, the credibility of results depends on the use of different plugins according to the nature of the experiment.

Among different plugins available in the QIIME2 software package, denoising is one of the most important and time-consuming steps, which can determine the outcomes of microbial community analysis. Denoising in computational bioinformatics is used to remove amplicon sequencing errors and aid to identify true biological signals (Callahan et al., 2016). Previous studies showed that denoising increases the accuracy of alpha diversity analysis compared to filtering data without using denoising (Reeder and Knight, 2010). The denoising step includes quality filtering, removal of chimera, and joining the paired-end reads (Callahan et al., 2016; Hall and Beiko, 2018; Estaki et al., 2020). DADA2 and Deblur are the two plugins used in QIIME2 to remove low-quality sequencing reads (sequencing errors, barcodes/tags, PCR errors, primer bias, sequencing depth bias, and chimeras) based on algorithms (Callahan et al., 2016; Amir et al., 2017). DADA2 denoiser develops amplicon sequence variants (ASVs) as their exact sequence variant, in contrast, deblur name these exact sequence variants as sub-operational taxonomic units (sOTUs) (Rosen et al., 2012; Callahan et al., 2017). ASVs develop in DADA2 describe features (taxa) in the microbial world with higher accuracy compared to sOTUs generated by clustering the same features into one group (Callahan et al., 2017). For example, one of the studies conducted to compare bioinformatics pipelines (DADA2, QIIME2-Deblur, and USEARCH-UNOISE3) reported the highest sensitivity in DADA2 compared to other denoising methods (Prodan et al., 2020). The authors reported that QIIME-uclust produced a large number of false positive OTUs and incorrect alpha diversity measures (Prodan et al., 2020). In addition, authors identified USEARCH-UNOISE3 shows the best balance between resolution and specificity compared to DADA2 and QIIME2-Deblur (Prodan et al., 2020). These studies suggest the importance of

careful selection and proper handling of computational bioinformatics tools for profiling microbial communities after sequencing.

Analytical biases generate during data analysis are critical in studying microbiota. Different bioinformatics pipelines have different plugins and algorithms that can use to filter sequencing errors and artificial sequences generated (Siegwald et al., 2019). These quality-filtering approaches mainly alter the outcomes of microbial profiling. Current bioinformatics platforms, however, incorporate error correction algorithms up to single nucleotide levels that can identify true biological insight. For example, DADA2 and Deblur are integrated into the QIIME2 pipeline, which allows accurately profile microbial communities (Hall and Beiko, 2018), if handle them properly based on the nature of experiment.

In summary, microbial profiling depends on the use of plugins available in a bioinformatics pipeline (Golob et al., 2017). Thus, subtle changes to parameters used in plugins can alter the outcomes of microbiota profiling. Use of default setting may result in false positive outcomes, which mislead the understanding of microbial community in an ecosystem. Hence, we suggest that customization/optimization of computational bioinformatics tools is crucial for the generation of credible information in microbial studies.

1.5 Knowledge gaps, hypothesis and objectives

The rapid advances in culture-independent microbial profiling techniques have improved the knowledge of microbiota. However, profiling of microbiota in low microbial biomass samples has become controversial due to the potential contaminations occur throughout the sequencing process. Although this can be controlled partly by the inclusion of appropriate controls, customization/optimization of bioinformatics pipelines will greatly enhance the outcomes of microbial profiling. In addition, the knowledge on the selection of genetic materials and the hypervariable region of the 16S rRNA gene to profile oral and fecal microbial communities in newborn beef calves using swabs samples (low microbial biomass samples) is limited.

This study hypothesized that the credibility/accuracy of microbial community composition generated by low microbial biomass samples depends on the use of proper molecular biological techniques, choice of genetic materials, choice of the hypervariable region of the 16S rRNA gene, and optimization of bioinformatics pipelines. The objectives of this thesis were to:

(1) optimize the denoising parameters in the QIIME2 bioinformatics pipeline to increase the accuracy of microbial community profiling in low microbial biomass samples (Chapter 2);

(2) assess the impact of genetic materials (DNA vs. RNA) on microbial profiles generated from low microbial biomass samples (Chapter 3),

(3) assess the impact of hypervariable regions of the 16S rRNA gene (V1V3 vs. V3V4) on oral and fecal microbial communities of newborn beef calves (Chapter 3).

1.6 References

Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. Sci. transl. med. 2014 May 21;6(237):237ra65-.

Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, et al. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. Msphere. 2021;6(1):e01202-20.

Al Nabhani Z, Eberl G. Imprinting of the immune system by the microbiota early in life. Mucosal Immunol. 2020;13(2):183-9.

Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2017;2(2).

Angly FE, Dennis PG, Skarshewski A, Vanwonterghem I, Hugenholtz P, Tyson GW. CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. Microbiome. 2014 ;2(1):1-3.

Araujo G, Yunta C, Terré M, Mereu A, Ipharraguerre I, Bach A. Intestinal permeability and incidence of diarrhea in newborn calves. J. dairy sci. 2015 ;98(10):7309-17.

Ballenghien M, Faivre N, Galtier N. Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. BMC Biol. 2017;15(1):25.

Ballou AL, Ali RA, Mendoza MA, Ellis JC, Hassan HM, Croom WJ, et al. Development of the Chick Microbiome: How Early Exposure Influences Future Microbial Diversity. Front Vet Sci. 2016;3:2.

Bäumler AJ, Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. Nature. 2016 ;535(7610):85-93.

Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform. 2021;22(1):178-93.

Bischoff SC. 'Gut health': a new objective in medicine?. BMC medicine. 2011 ;9:1-4.

Bleidorn C, Bleidorn C. Sequencing techniques. Phylogenomics: An Introduction. 2017:43-60.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852-7.

Caballero-Flores G, Pickard JM, Nunez G. Microbiota-mediated colonization resistance: mechanisms and regulation. Nat Rev Microbiol. 2022.

Caballero-Flores G, Pickard JM, Núñez G. Microbiota-mediated colonization resistance: Mechanisms and regulation. Nat. Rev. Microbiol. 2023 ;21(6):347-60.

Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11(12):2639-43.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.

Carlsen T, Aas AB, Lindner D, Vrålstad T, Schumacher T, Kauserud H. Don't make a mista (g) ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies?. Fungal Ecol. 2012;5(6):747-9.

Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. microbiol. methods. 2007;69(2):330-9.

Chen C, Khaleel SS, Huang H, Wu CH. Software for pre-processing Illumina next-generation sequencing short read sequences. Source code biol. med. 2014;9(1):1-1.

Collado MC, Cernada M, Neu J, Perez-Martinez G, Gormaz M, Vento M. Factors influencing gastrointestinal tract and microbiota immune interaction in preterm infants. Pediatr Res. 2015;77(6):726-31.

Collado MC, Rautava S, Aakko J, Isolauri E, Salminen S. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. Sci Rep. 2016;6:23129.

Constantinides MG, Link VM, Tamoutounour S, Wong AC, Perez-Chaparro PJ, Han SJ, et al. MAIT cells are imprinted by the microbiota in early life and promote tissue repair. Science. 2019;366(6464).

Correa-Fiz F, Goncalves Dos Santos JM, Illas F, Aragon V. Antimicrobial removal on piglets promotes health and higher bacterial diversity in the nasal microbiota. Sci Rep. 2019;9(1):6545.

Crosby CM, Kronenberg M. Tissue-specific functions of invariant natural killer T cells. Nat. Rev. Immunol. 2018 ;18(9):559-74.

Danzeisen JL, Clayton JB, Huang H, Knights D, McComb B, Hayer SS, et al. Temporal Relationships Exist Between Cecum, Ileum, and Litter Bacterial Microbiomes in a Commercial Turkey Flock, and Subtherapeutic Penicillin Treatment Impacts Ileum Bacterial Community Establishment. Front Vet Sci. 2015;2:56.

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(1):226.

De Vrieze J, Pinto AJ, Sloan WT, Ijaz UZ. The active microbial community more accurately reflects the anaerobic digestion process: 16S rRNA (gene) sequencing as a predictive tool. Microbiome. 2018;6(1):63.

Doyle RM, Alber DG, Jones HE, Harris K, Fitzgerald F, Peebles D, et al. Term and preterm labour are associated with distinct microbial community structures in placental membranes which are independent of mode of delivery. Placenta. 2014;35(12):1099-101.

Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol. 2019;27(2):105-17.

Ellermann M, Carr JS, Fodor AA, Arthur JC, Carroll IM. Characterizing and functionally defining the gut microbiota: methodology and implications. InThe microbiota in gastrointestinal pathophysiology 2017 (pp. 15-25). Academic Press.

Estaki M, Jiang L, Bokulich NA, McDonald D, Gonzalez A, Kosciolek T, et al. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. Curr Protoc Bioinformatics. 2020;70(1):e100.

Fischer AJ, Song Y, He Z, Haines DM, Guan LL, Steele MA. Effect of delaying colostrum feeding on passive transfer and intestinal bacterial colonization in neonatal male Holstein calves. J Dairy Sci. 2018;101(4):3099-109.

Fouhse JM, Yang K, More-Bayona J, Gao Y, Goruk S, Plastow G, et al. Neonatal Exposure to Amoxicillin Alters Long-Term Immune Response Despite Transient Effects on Gut-Microbiota in Piglets. Front Immunol. 2019;10:2059.

Fraher MH, O'Toole PW, Quigley EM. Techniques used to characterize the gut microbiota: a guide for the clinician. Nat Rev Gastroenterol Hepatol. 2012;9(6):312-22.

Furman O, Shenhav L, Sasson G, Kokou F, Honig H, Jacoby S, Hertz T, Cordero OX, Halperin E, Mizrahi I. Stochasticity constrained by deterministic effects of diet and age drive rumen microbiome assembly dynamics. Nat. commun. 2020 ;11(1):1904.

Geleto A, Chojenta C, Taddele T, Loxton D. Association between maternal mortality and caesarean section in Ethiopia: a national cross-sectional study. BMC Pregnancy and childbirth. 2020 ;20:1-1.

Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. Science. 2016 ;352(6285):539-44.

Ghaffari MH, Sadri H, Steinhoff-Wagner J, Hammon HM, Sauerwein H. Effects of colostrum feeding on the mRNA abundance of genes related to toll-like receptors, key antimicrobial defense molecules, and tight junctions in the small intestine of neonatal dairy calves. J Dairy Sci. 2021;104(9):10363-73.

Ghosh S, Whitley CS, Haribabu B, Jala VR. Regulation of Intestinal Barrier Function by Microbial Metabolites. Cell Mol Gastroenterol Hepatol. 2021;11(5):1463-82.

Gillilland GM, Young VB, Huffnagle GB. Gastrointestinal Microbial Ecology with Perspectives on Health and Disease. Editor(s): Leonard R. Johnson, Fayez K. Ghishan, Jonathan D. Kaunitz, Juanita L. Merchant, Hamid M. Said, Jackie D. Wood, Physiology of the Gastrointestinal Tract (Fifth Edition) 2012 (pp 1119-1134) Academic Press.

Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog. 2016;8:24.

Golob JL, Margolis E, Hoffman NG, Fredricks DN. Evaluating the accuracy of amplicon-based microbiome computational pipelines on simulated human gut microbial communities. BMC Bioinformatics. 2017;18(1):283.

Gonzalez-Gonzalez M, Diaz-Zepeda C, Eyzaguirre-Velasquez J, Gonzalez-Arancibia C, Bravo JA, Julio-Pieper M. Investigating Gut Permeability in Animal Models of Disease. Front Physiol. 2018;9:1962.

Guevarra RB, Lee JH, Lee SH, Seok MJ, Kim DW, Kang BN, et al. Piglet gut microbial shifts early in life: causes and effects. J Anim Sci Biotechnol. 2019;10:1.

Guzman CE, Wood JL, Egidi E, White-Monsant AC, Semenec L, Grommen SVH, et al. A pioneer calf foetus microbiome. Sci Rep. 2020;10(1):17712.

Hall M, Beiko RG. 16S rRNA gene analysis with QIIME2. Microbiome analysis: methods and protocols. 2018:113-29.

Hansson J, Bosco N, Favre L, Raymond F, Oliveira M, Metairon S, Mansourian R, Blum S, Kussmann M, Benyacoub J. Influence of gut microbiota on mouse B2 B cell ontogeny and function. Mol. immunol. 2011;48(9-10):1091-101.

Ho NT, Li F, Lee-Sarwar KA, Tun HM, Brown BP, Pannaraj PS, Bender JM, Azad MB, Thompson AL, Weiss ST, Azcarate-Peril MA. Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. Nat. commun. 2018;9(1):4169.

Hornef MW, Torow N. 'Layered immunity'and the 'neonatal window of opportunity'–timed succession of non-redundant phases to establish mucosal host–microbial homeostasis after birth. Immunology. 2020;159(1):15-25.

Jamuar SS, D'Gama AM, Walsh CA. Somatic mosaicism and neurological diseases. InGenomics, circuits, and pathways in clinical neuropsychiatry 2016 (pp. 179-199). Academic Press.

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):5029.

Jovel J, Nimaga A, Jordan T, O'Keefe S, Patterson J, Thiesen A, et al. Metagenomics Versus Metatranscriptomics of the Murine Gut Microbiome for Assessing Microbial Metabolism During Inflammation. Front Microbiol. 2022;13:829378.

Kameoka S, Motooka D, Watanabe S, Kubo R, Jung N, Midorikawa Y, et al. Benchmark of 16S rRNA gene amplicon sequencing using Japanese gut microbiome data from the V1-V2 and V3-V4 primer sets. BMC Genomics. 2021;22(1):527.

Karstens L, Asquith M, Caruso V, Rosenbaum JT, Fair DA, Braun J, et al. Community profiling of the urinary microbiota: considerations for low-biomass samples. Nat Rev Urol. 2018;15(12):735-49.

Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput Biol. 2012;8(10):e1002743.

Kennedy KM, de Goffau MC, Perez-Munoz ME, Arrieta MC, Backhed F, Bork P, et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. Nature. 2023;613(7945):639-49.

Kennedy KM, Gerlach MJ, Adam T, Heimesaat MM, Rossi L, Surette MG, et al. Fetal meconium does not have a detectable microbiota before birth. Nat Microbiol. 2021;6(7):865-73.

Kim ET, Lee SJ, Kim TY, Lee HG, Atikur RM, Gu BH, et al. Dynamic Changes in Fecal Microbial Communities of Neonatal Dairy Calves by Aging and Diarrhea. Animals (Basel). 2021;11(4).

Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nature Reviews Microbiology. 2018;16(7):410-22.

Kogut MH, Arsenault RJ. Editorial: Gut Health: The New Paradigm in Food Animal Production. Front Vet Sci. 2016;3:71.

Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev. 2015;28(1):237-64.

Lamouse-Smith ES, Tzeng A, Starnbach MN. The intestinal flora is required to support antibody responses to systemic immunization in infant and germ free mice. PLoS One. 2011;6(11):e27662.

Larsson AJM, Stanley G, Sinha R, Weissman IL, Sandberg R. Computational correction of index switching in multiplexed sequencing libraries. Nat Methods. 2018;15(5):305-7.

Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, et al. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. Microbiome. 2016;4(1):29.

Li F, Henderson G, Sun X, Cox F, Janssen PH, Guan le L. Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches. Front Microbiol. 2016;7:987.

Li N, Huang S, Jiang L, Wang W, Li T, Zuo B, et al. Differences in the Gut Microbiota Establishment and Metabolome Characteristics Between Low- and Normal-Birth-Weight Piglets During Early-Life. Front Microbiol. 2018;9:1798.

Li Y, Xia S, Jiang X, Feng C, Gong S, Ma J, et al. Gut Microbiota and Diarrhea: An Updated Review. Front Cell Infect Microbiol. 2021;11:625210.

Lim ES, Rodriguez C, Holtz LR. Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. Microbiome. 2018;6(1):87.

Litvak Y, Byndloss MX, Bäumler AJ. Colonocyte metabolism shapes the gut microbiota. Science. 2018;362(6418):eaat9076.

Loh L, Ivarsson MA, Michaëlsson J, Sandberg JK, Nixon DF. Invariant natural killer T cells developing in the human fetus accumulate and mature in the small intestine. Mucosal immunol. 2014;7(5):1233-43.

Louca S, Doebeli M, Parfrey LW. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome. 2018;6(1):41.

Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One. 2012;7(2):e30087.2018;9:2755.

Ma J, Li Z, Zhang W, Zhang C, Zhang Y, Mei H, et al. Comparison of gut microbiota in exclusively breast-fed and formula-fed babies: a study of 91 term infants. Sci Rep. 2020;10(1):15792.

Ma T, O'Hara E, Song Y, Fischer AJ, He Z, Steele MA, et al. Altered mucosa-associated microbiota in the ileum and colon of neonatal calves in response to delayed first colostrum feeding. J Dairy Sci. 2019;102(8):7073-86.

Ma T, Villot C, Renaud D, Skidmore A, Chevaux E, Steele M, et al. Linking perturbations to temporal changes in diversity, stability, and compositions of neonatal calf gut microbiota: prediction of diarrhea. ISME J. 2020;14(9):2223-35.

Malmuthuge N, Griebel PJ, Guan le L. The Gut Microbiome and Its Potential Role in the Development and Function of Newborn Calf Gastrointestinal Tract. Front Vet Sci. 2015;2:36.

Malmuthuge N, Griebel PJ. Fetal environment and fetal intestine are sterile during the third trimester of pregnancy. Vet Immunol Immunopathol. 2018;204:59-64.

Malmuthuge N, Guan LL. Understanding the gut microbiome of dairy calves: Opportunities to improve early-life gut health. J Dairy Sci. 2017;100(7):5996-6005.

Malmuthuge N, Liang G, Guan LL. Regulation of rumen development in neonatal ruminants through microbial metagenomes and host transcriptomes. Genome Biol. 2019;20(1):172.

Mardis ER. DNA sequencing technologies: 2006–2016. Nat. protoc. 2017 ;12(2):213-8.

Marizzoni M, Gurry T, Provasi S, Greub G, Lopizzo N, Ribaldi F, et al. Comparison of Bioinformatics Pipelines and Operating Systems for the Analyses of 16S rRNA Gene Amplicon Sequences in Human Fecal Samples. Front Microbiol. 2020;11:1262.

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(1):35.

Maynard CL, Elson CO, Hatton RD, Weaver CT. Reciprocal interactions of the intestinal microbiota and immune system. Nature. 2012 ;489(7415):231-41.

McElroy SJ, Prince LS, Weitkamp JH, Reese J, Slaughter JC, Polk DB. Tumor necrosis factor receptor 1-dependent depletion of mucus in immature small intestine: a potential role in neonatal necrotizing enterocolitis. Am J Physiol Gastrointest Liver Physiol. 2011;301(4):G656-66.

McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217.

Minich JJ, Zhu Q, Janssen S, Hendrickson R, Amir A, Vetter R, Hyde J, Doty MM, Stillwell K, Benardini J, Kim JH. KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. MSystems. 2018 ;3(3):e00218-17.

Mishra A, Lai GC, Yao LJ, Aung TT, Shental N, Rotter-Maskowitz A, et al. Microbial exposure during early human development primes fetal immune cells. Cell. 2021;184(13):3394-409 e20.

Mohsen A, Park J, Chen YA, Kawashima H, Mizuguchi K. Impact of quality trimming on the efficiency of reads joining and diversity analysis of Illumina paired-end reads in the context of QIIME1 and QIIME2 microbiome analysis frameworks. BMC Bioinformatics. 2019;20(1):581.

Motley ST, Picuri JM, Crowder CD, Minich JJ, Hofstadler SA, Eshoo MW. Improved multiple displacement amplification (iMDA) and ultraclean reagents. BMC genomics. 2014 ;15:1-0.

Mulder IE, Schmidt B, Stokes CR, Lewis M, Bailey M, Aminov RI, et al. Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. BMC Biol. 2009;7:79.

Munshi A, Sharma V. Omics and edible vaccines. InOmics Technologies and Bio-Engineering 2018 (pp. 129-141). Academic Press.

Nagpal R, Tsuji H, Takahashi T, Kawashima K, Nagata S, Nomoto K, et al. Sensitive Quantitative Analysis of the Meconium Bacterial Microbiota in Healthy Term Infants Born Vaginally or by Cesarean Section. Front Microbiol. 2016;7:1997.

Nakandalage R, Guan L L, Malmuthuge N, et al. Microbial Interventions to Improve Neonatal Gut Health. Microorganisms. 2023*;* 11, no. 5: 1328.

Nguyen NH, Smith D, Peay K, Kennedy P. Parsing ecological signal from noise in next generation amplicon sequencing. New Phytologist. 2015 ;205(4):1389-93.

Nilsson RH, Ryberg M, Abarenkov K, Sjokvist E, Kristiansson E. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. FEMS Microbiol Lett. 2009;296(1):97-101.

Nuzhat S, Hasan SMT, Palit P, Islam MR, Mahfuz M, Islam MM, et al. Effects of probiotic and synbiotic supplementation on ponderal and linear growth in severely malnourished young infants in a randomized clinical trial. Sci Rep. 2023;13(1):1845.

Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, et al. Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function. 2012;336(6080):489-93.

Pluske JR, Turpin DL, Kim JC. Gastrointestinal tract (gut) health in the young pig. Anim Nutr. 2018;4(2):187-96.

Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS One. 2020;15(1):e0227434.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat. biotechnol. 2017;35(9):833-44.

Rackaityte E, Halkias J, Fukui EM, Mendoza VF, Hayzelden C, Crawford ED, et al. Viable bacterial colonization is highly limited in the human intestine in utero. Nat Med. 2020;26(4):599-607.

Radhakrishnan ST, Gallagher KI, Mullish BH, Serrano-Contreras JI, Alexander JL, Miguens Blanco J, Danckert NP, Valdivia-Garcia M, Hopkins BJ, Ghai A, Ayub A. Rectal swabs as a viable alternative to faecal sampling for the analysis of gut microbiota functionality and composition. Sci Rep. 2023 ;13(1):493.

Rai SN, Qian C, Pan J, Rai JP, Song M, Bagaitkar J, et al. Microbiome data analysis with applications to pre-clinical studies using QIIME2: Statistical considerations. Genes Dis. 2021;8(2):215-23.

Raman AS, Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Subramanian S, et al. A sparse covarying unit that describes healthy and impaired human gut microbiota development. Science. 2019;365(6449).

Ravisankar S, Tatum R, Garg PM, Herco M, Shekhawat PS, Chen YH. Necrotizing enterocolitis leads to disruption of tight junctions and increase in gut permeability in a mouse model. BMC Pediatr. 2018;18(1):372.

Razali H, O'Connor E, Drews A, Burke T, Westerdahl H. A quantitative and qualitative comparison of illumina MiSeq and 454 amplicon sequencing for genotyping the highly polymorphic major histocompatibility complex (MHC) in a non-model species. BMC Res Notes. 2017;10(1):346.

Reeder J, Knight R. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Methods. 2010;7(9):668-9.

Reyman M, van Houten MA, van Baarle D, Bosch AA, Man WH, Chu ML, Arp K, Watson RL, Sanders EA, Fuentes S, Bogaert D. Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. Nat. commun. 2019 ;10(1):4997.

Rosa F, Michelotti TC, St-Pierre B, Trevisi E, Osorio JS. Early life fecal microbiota transplantation in neonatal dairy calves promotes growth performance and alleviates inflammation and oxidative stress during weaning. Animals. 2021 ;11(9):2704.

Rosen MJ, Callahan BJ, Fisher DS, Holmes SP. Denoising PCR-amplified metagenome data. BMC bioinformatics. 2012 ;13(1):1-6.

Rothberg JM, Leamon JH. The development and impact of 454 sequencing. Nat Biotechnol. 2008;26(10):1117-24.

Saffarian A, Mulet C, Regnault B, Amiot A, Tran-Van-Nhieu J, Ravel J, Sobhani I, Sansonetti PJ, Pédron T. Crypt-and mucosa-associated core microbiotas in humans and their alteration in colon cancer patients. MBio. 2019 ;10(4):e01315-19.

Salgar-Chaparro SJ, Machuca LL. Complementary DNA/RNA-Based Profiling: Characterization of Corrosive Microbial Communities and Their Functional Profiles in an Oil Production Facility. Front Microbiol. 2019;10:2587.

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC biology. 2014 ;12:1-2.

Sangild PT, Strunk T, Currie AJ, Nguyen DN. Editorial: Immunity in Compromised Newborns. Front Immunol. 2021;12:732332.

Satokari R, Gronroos T, Laitinen K, Salminen S, Isolauri E. Bifidobacterium and Lactobacillus DNA in the human placenta. Lett Appl Microbiol. 2009;48(1):8-12.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41.

Schokker D, Zhang J, Zhang LL, Vastenhouw SA, Heilig HG, Smidt H, et al. Early-life environmental variation affects intestinal microbiota and immune development in new-born piglets. PLoS One. 2014;9(6):e100040.

Seitz V, Schaper S, Droge A, Lenze D, Hummel M, Hennig S. A new method to prevent carry-over contaminations in two-step PCR NGS library preparations. Nucleic Acids Res. 2015;43(20):e135.

Siegwald L, Caboche S, Even G, Viscogliosi E, Audebert C, Chabe M. The Impact of Bioinformatics Pipelines on Microbiota Studies: Does the Analytical "Microscope" Affect the Biological Interpretation? Microorganisms. 2019;7(10).

Slanzon GS, Ridenhour BJ, Parrish LM, Trombetta SC, Moore DA, Sischo WM, et al. Effects of a farm-specific fecal microbial transplant (FMT) product on clinical outcomes and fecal microbiome composition in preweaned dairy calves. PLoS One. 2022;17(10):e0276638.

Song SJ, Wang J, Martino C, Jiang L, Thompson WK, Shenhav L, et al. Naturalization of the microbiota developmental trajectory of Cesarean-born neonates after vaginal seeding. Med (N Y). 2021;2(8):951-64 e5.

Song SJ, Wang J, Martino C, Jiang L, Thompson WK, Shenhav L, et al. Naturalization of the microbiota developmental trajectory of Cesarean-born neonates after vaginal seeding. Med (N Y). 2021;2(8):951-64 e5.

Stinson L, Hallingstrom M, Barman M, Viklund F, Keelan J, Kacerovsky M, et al. Comparison of Bacterial DNA Profiles in Mid-Trimester Amniotic Fluid Samples From Preterm and Term Deliveries. Front Microbiol. 2020;11:415.

Strathdee F, Free A. Denaturing gradient gel electrophoresis (DGGE). DNA electrophoresis: Methods Protoc. 2013:145-57.

Swanson KS, Gibson GR, Hutkins R, Reimer RA, Reid G, Verbeke K, et al. The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of synbiotics. Nat Rev Gastroenterol Hepatol. 2020;17(11):687-701.

Thomson CA, Morgan SC, Ohland C, McCoy KD. From germ-free to wild: modulating microbiome complexity to understand mucosal immunology. Mucosal Immunol. 2022;15(6):1085-94.

Vogt SL, Finlay BB. Gut microbiota-mediated protection against diarrheal infections. J Travel Med. 2017;24(suppl_1):S39-S43.

Wang L, Cao H, Liu L, Wang B, Walker WA, Acra SA, et al. Activation of epidermal growth factor receptor mediates mucin production stimulated by p40, a Lactobacillus rhamnosus GG-derived protein. 2014;289(29):20234-44.

Wang X, Jordan IK, Mayer LW. A phylogenetic perspective on molecular epidemiology. Mol. microbiol. 2015 Jan 1:517-36.

Wang Y, Gong L, Wu YP, Cui ZW, Wang YQ, Huang Y, et al. Oral administration of Lactobacillus rhamnosus GG to newborn piglets augments gut barrier function in pre-weaning piglets. J Zhejiang Univ Sci B. 2019;20(2):180-92.

Weyrich LS, Duchene S, Soubrier J, Arriola L, Llamas B, Breen J, et al. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. Nature. 2017;544(7650):357-61.

Wu Y, Nie C, Luo R, Qi F, Bai X, Chen H, Niu J, Chen C, Zhang W. Effects of multispecies probiotic on intestinal microbiota and mucosal barrier function of neonatal calves infected with E. coli K99. Front. Microbiol. 2022:3933.

Xiang Q, Yan X, Shi W, Li H, Zhou K. Early gut microbiota intervention in premature infants: Application perspectives. J Adv Res. 2022.

Yamamoto, Masahiro, Rui Yamaguchi, Kaori Munakata, Kiyoe Takashima, Mitsue Nishiyama, Kyoji Hioki, Yasuyuki Ohnishi et al. A microarray analysis of gnotobiotic mice indicating that microbial exposure during the neonatal period plays an essential role in immune system development. BMC genomics 13, no. 1 (2012): 1-15.

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486(7402):222-7.

Zhang X, Zhivaki D, Lo-Man R. Unique aspects of the perinatal immune system. Nat Rev Immunol. 2017;17(8):495-507.

Zhou P, Chen R, Zou CY, Zhang M, Chen ZH, Ma LY. Effect of perinatal factors on postnatal development of lymphocyte subsets in early preterm infants. Chin. J. Contemp. Pediatr. 2016 ;18(10):953-9.

Zhu L, Luo F, Hu W, Han Y, Wang Y, Zheng H, et al. Bacterial Communities in the Womb During Healthy Pregnancy. Front Microbiol. 2018;9:2163.

Zitvogel L, Kroemer G. Commensals shape the immune system. Nat. Rev. Immunol. 2021 ;21(10):615-.

1.7 Tables and Figures

Table 1.1 Limitations of microbial community profiling using low microbial biomass samples.

Limitation	References
Contaminations during sampling and laboratory environment	Nguyen, et al., 2015 Eisenhofer et al., 2019
Contamination from plastic consumables	Motley et al., 2014
Contaminations from researchers/technicians	Eisenhofer *et al., 2019*
Contaminations from nucleic acid extraction kits	Weyrich et al., 2017; Salter et al., 2014; Lauder et al., 2016; Glassing et al., 2016
Contaminations from PCR master mix	Eisenhofer *et al., 2019*
Cross contaminations from other samples	Seitz et al., 2015; Ballenghien et al., 2017
Cross contamination during sequencing run	Seitz et al., 2015; Ballenghien et al., 2017
Use of default settings/lack of customization of bioinformatics tools	Eisenhofer *et al.,* 2019

Table 1.2 Advantages and disadvantages of different molecular techniques to profile microbiota.

Technique	Advantages	Disadvantages	References
Denaturing Gradient Gel Electrophoresis (DGGE)	Short to medium-length DNA fragments for further analysis, fast	Less reproducibility, labor intensiveness, limited sequencing information	Fraher et al., 2012; Strathdee and Free, 2013
Terminal Restriction Fragment Length Polymorphism (T-RFLP)	Quantitative comparison through visualization of DNA fragments in a gel by fluorescent labeling	Less reproducibility, labor intensiveness, limited sequencing information	Fraher et al., 2012
Fluorescence *In Situ* Hybridization (FISH)	Fluorescent-labeled oligonucleotide probes to target bacterial DNA sequences, phylogenetic identification	Unable to identify unknown species (probe designed to target specific bacterial species)	Fraher et al., 2012
DNA microarrays	Phylogenetic identification, fast	Limited detection in low abundance species	Fraher et al., 2012
454 pyrosequencing	Phylogenetic identification, fast, identify unknown species	High cost, high error rate	Rothberg and Leamon, 2008; Bleidorn, 2017
Illumina sequencing	High throughput read counts with lower cost and error rates	Biases through PCR and sequencing	Luo et al., 2012
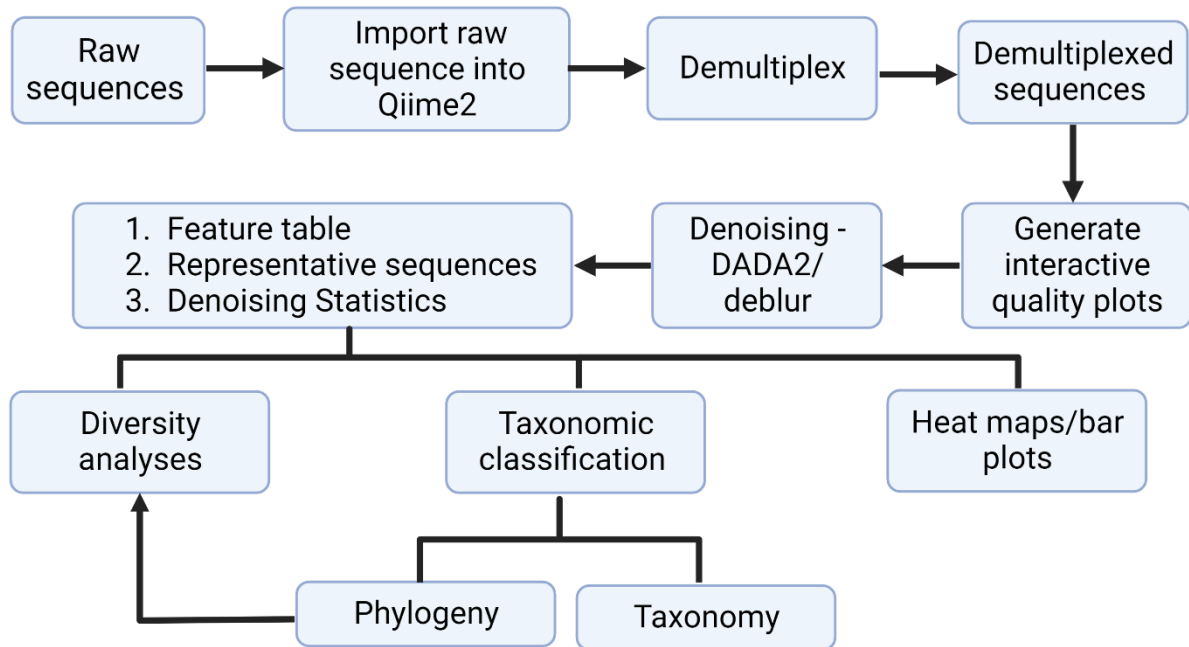
Figure 1.1 Basic workflow of Quantitative Insights into Microbial Ecology (QIIME2; https://qiime2.org)

bioinformatics pipeline.

Chapter 2. Customization of denoising parameters in QIIME2 bioinformatics platform to increase the accuracy of microbial community profiling

2.1 Abstract

Rapid advancements in sequencing and user-friendly bioinformatics platforms have improved the knowledge of microbial community composition. However, the credibility of the data generated through next-generation sequencing (NGS) and computational bioinformatics tools has become debatable. The use of appropriate approaches to control contaminations during sample processing and sequencing and the optimization of bioinformatics pipelines determine the accuracy of NGS-based microbial profiling. QIIME2 is one of the most used bioinformatics pipelines that allows users to perform quality filtering, classification, community analysis, visualization, and statistical analysis through one open-source software package. Denoising in QIIME2 is one of the important plugins for quality filtering, which should be handled carefully to generate credible outcomes from microbial community profiling. This study aimed to optimize the denoising parameters to increase the accuracy of microbial community data analysis when using biological samples containing limited microbial materials (low-microbial biomass samples). This study used the V1V3 region of the 16S rRNA gene to profile the fecal microbial community of newborn beef calves using rectal swabs and data were analyzed using QIIME2. Use of optimized (truncation: forward – 294; reverse – 241; median quality score - >25) denoising parameters increased the percentage of merged read (default – 1%; optimized – 45%), and the number of samples used for downstream analysis compared to default (truncation: forward – 281; reverse – 207; mean quality score - >25) denoising parameters. Moreover, the optimization of denoising parameters improved microbial diversity and taxonomic classification. Our findings revealed that the default settings of the bioinformatics tools are not suitable for all microbial analyses. The customization of plugins in bioinformatics pipelines needs to be considered to obtain credible outcomes in microbial community assessments.

Keywords: bioinformatics tools, denoising, microbial community composition, optimization

2.2 Introduction

The identification of early-life microbial colonization in humans and other animals has gained more attention due to its influence on overall host health (Kogut et al., 2016; Pluske et al., 2018; Nakandalage et al., 2023). Profiling of microbial communities has become highly available to researchers due to the advanced next-generation sequencing technologies (NGS) and user-friendly microbial analysis platforms (Callahan et al., 2017). A number of methods (Mothur, Quantitative Insights into Microbial Ecology (QIIME2;https://qiime2.org), and phyloseq) have been introduced to perform microbial data analysis (Schloss et al., 2009; Reeder and Knight, 2010; McMurdie and Holmes, 2013; Bolyen et al., 2019). For example, Mothur is an open-source software package used to perform data analysis, however, it requires other software (eg: R statistical software package) to visualize data (Bolyen et al., 2019). Among these microbiome data analysis software, QIIME2 provides researchers with one open-source software package to conduct data analysis, visualization, and statistical analysis (Bolyen et al., 2019; Mohsen et al., 2019). Thus, QIIME2 has been widely used for microbial data analysis.

QIIME2 contains a series of plugins (demultiplex (demux), denoising (dada2/deblur), generation of the feature table, and taxonomic classification (feature-classifier), and diversity analysis useful for analyzing raw sequencing reads and generate readable graphical outcomes (Mohsen et al., 2019). QIIME2 also provides the opportunity to combine different tools (e.g., cutadapt) and different options (e.g., DADA2, deblur) to perform the same task (Martin, 2011; Callahan et al., 2016; Amir et al., 2017; Telatin, 2021). However, the credibility of data generated through bioinformatics platforms depends on the use of appropriate plugins and parameters according to the experiment nature. For example, Bokulich and colleagues (2018), introduced q2-feature classifier plugin for microbial data analysis after optimizing previously available Qiime1 taxonomic classification methods. They optimized the commonly used taxonomic classification methods in Qiime1 (e.g., uclust, BLAST) to increase the species-level classification by removing ambiguous and null reference sequences. They also introduced QIIME2 q2-feature classifiers (e.g., naïve-Bayes, BLAST+, VSERACH) to accurately classify marker gene-based amplicon sequencing data. In addition, they also mentioned that each classifier required some degree of optimization (Bokulich et al., 2018), suggesting the importance of optimization of bioinformatics plugins in microbial data analysis. However, the optimization of one condition may not translate to another (Bokulich et al., 2018).

Denoising is one of the important and time-consuming plugins in QIIME2, which removes amplicon sequencing errors and chimera sequences and joins the paired-end reads (Callahan et al., 2016; Hall and Beiko, 2018; Estaki et al., 2020). For example, one of the previous studies showed that the denoising tool increases the accuracy of alpha diversity analysis compared to quality filtering without the denoising step (Reeder and Knight, 2010). DADA2 and Deblur are two different plugins used in QIIME2 for denoising (Callahan et al., 2016; Amir et al., 2017). A comparison of different denoising plugins in a previous study reported that higher sensitivity in DADA2 denoiser compared to other denoising tools such as DADA2-debluar and USERACH-UNOISE3 (Prodan et al., 2020). Moreover, DADA2 denoiser generates amplicon sequence variants (ASVs) as their exact sequence variant and describes microbial taxa with higher accuracy compared to sub-operational taxonomic units generated by Deblur (Rosen et al., 2012; Callahan et al., 2017).

The use of default settings in denoising might lead to losing information on microbial community composition due to the lower merge reads after denoising, suggesting default settings in the DADA2 denoising plugin might not be applicable for all microbiome analyses. The objective of this study is to optimize the denoising parameters to increase the accuracy of microbial community profiling in low microbial biomass samples.

2.3 Materials and Methods

2.3.1   Animal experiments and sample collection

All experiment protocols were approved by the Livestock Care Committee at the University of Alberta (AUP-00004183) and the University of Saskatchewan (AUP-20170015) and were conducted following the guidelines of the Canadian Council on Animal Care. Newborn beef calves (n=20) were obtained from the Rayner Dairy Research and Teaching Facility, University of Saskatchewan (Saskatoon, SK) within 24 hours of birth (Figure 2.1). Rectal swabs of newborn beef calves were collected after birth prior to colostrum feeding (Day 1-D1). Calves were allowed to suckle colostrum from cows and re-sample after 24 hours (Day 2-D2). All samples were snap frozen in liquid nitrogen and stored at -80°C for nucleic acid extraction.

2.3.2   Nucleic acid extraction

Total genomic DNA from rectal swabs were extracted using the QIAamp Fast DNA stool Mini kit (Qiagen, USA), with a fast spin-column procedure. Briefly, swab samples were suspended in 1 mL InhibitEx buffer

and lysed protein using 25 µL of proteinase K. DNA was eluted using elution buffer after several cleaning steps following the manufacturer's instructions. The quality and quantity of the DNA were evaluated using the Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Delaware, USA).

2.3.3 Profiling of fecal bacterial communities in newborn beef calves using amplicon sequencing of the 16S rRNA gene

Diluted DNA (50ng/µL) was used to amplify V1V3 hypervariable regions of the 16S rRNA gene using universal bacterial primers 9F (5'-GAGTTTGATCMTGGCTCAG-3') and 515R (5'-CCGCGGCKGCTGGCAC-3') (Li et al., 2019). Amplicons were barcoded and sequenced on an Illumina Miseq 300 with paired-end reads at the McGill University and Genome Quebec Innovation Center (Quebec, Canada).

2.3.4 Optimization of denoising parameters in QIIME2 bioinformatics platform for microbial profiling

Raw sequence data were imported into QIIME2 software version 2022.2. Demultiplexed sequencing reads were assigned to samples using the q2-demux plugin in the QIIME2 interface and visualized the interactive quality plots (Figure 2.2). Based on the interactive quality plots, forward and reverse truncation lengths were selected as 281 bp and 207 bp, respectively to maintain a mean Phred quality score ≥25. Feature table, representative sequence and denoising statistics files were generated after denoising. Denoising statistics were reviewed to identify the number of merged reads counts after denoising before continuing to downstream analysis.

Following reviewing the denoising statics of default parameter-based denoising, they were optimized to maintain a minimum 20 bp long overlap between forward and reverse sequences as follows. Forward truncation length was increased from 281 to 294 and reverse truncation length from 207 to 241. Sequence quality to retain high-quality sequences was chosen as a median Phred score ≥25.

2.3.5 Downstream analysis to profile fecal bacterial community

Feature tables generated after denoising were subjected to three different quality filtering steps; 1. filtering low abundant ASVs (<0.005% of samples), 2. filtering environmental contamination (mitochondria, chloroplast, cyanobacteria, chloroflexi, archaea), and 3. filtering ASVs present in NTC to remove contaminated and artificial ASVs. Then, the remaining good-quality sequences were used to assign

taxonomy using the SILVA database. SILVA138-99 classifier was trained separately for V1V3 and V3V4 primers prior to taxonomic assignment. Diversity analysis was performed using the diversity plugin in QIIME2. Fecal microbial diversity analysis was performed using alpha diversity indices (Shannon, Chao1, Faith's phylogenetic diversity, Fisher's index, Simpson's index, Pielou's evenness, and Good's coverage) and beta diversity (Weighted and unweighted UniFrac Distance Matrix) within QIIME2 (2022.2 version). A relative abundance of identified taxonomies was generated at the phylum and genus levels.

2.3.6   Statistical Analysis

One-way ANOVA was conducted to compare sequencing statistics between default and optimized denoising parameters after testing for normality using a normal distribution test (PROC UNIVARIATE in SAS). Mean separation was carried out using Duncan's multiple-range test to declare significances at P < 0.05. Alpha diversity indices between default and optimized denoising parameters were compared using Kruskal Wallis non-parametric test. Weighted and unweighted UniFrac distance matrices were analyzed using PERMANOVA statistical test to compare microbial communities generated by two different methods.

2.4 Results

2.4.1   Impact of optimization of denoising parameters on sequencing statistics

Optimization of denoising parameters increased the merge read counts nearly by 60 times (default – 504±104; optimized – 29,312±468) and ASVs by 12 times (default – 26±5; optimized – 337±49) compared to default denoising parameters (Table 2.1). After optimization of denoising parameters, the percentage of input merge increased from 1% to 45%. Optimization of denoising parameters also increased the number of samples used for downstream analysis from sixteen to twenty and one to twenty on D1 and D2, respectively (Table 2.1). Optimization of denoising parameters increased the sequence count in each fecal sample compared to default denoising parameters (Table 2.2).

2.4.2   Impact of optimization of denoising parameters on fecal bacterial diversity

Optimization of denoising parameters increased (p< 0.0001) the Chao1 (richness) by five times, Faith's phylogenetic diversity by four times and Fisher's index by three times compared to default denoising parameters (Table 2.3). Optimization of denoising parameters slightly increased the Shannon (evenness and richness; p< 0.0001) and Simpson's index (p = 0.0009) of the fecal bacterial community compared to

default denoising parameters (Table 2.3). In contrast, similar values were observed for Pielou's evenness and Good's coverage.

Default denoising parameters only captured 46% and 34% of bacterial data based on weighted and unweighted UniFrac distance matrix, respectively (Figure 2.3 A). However, optimization of denoising parameters increased the data captured for downstream analysis up to 71% and 67% based on weighted and unweighted UniFrac distance matrix, respectively (Figure 2.3 B). An increase in the data representation with the optimized denoising parameters revealed that 53% (weighted UniFrac distance matrix) and 57% (unweighted UniFrac distance matrix) of the observed variations among bacterial communities could be explained by sampling time point (Figure 2.3B). In contrast, only 16% (weighted UniFrac distance matrix) and 8% (unweighted UniFrac distance matrix) of observed variation between bacterial communities could be explained by sampling time points when using default denoising parameters (Figure 2.3A).

2.4.3 Impact of optimization of denoising parameters on taxonomic assessment

Optimization of denoising parameters showed distinct taxonomic classification compared to default denoising parameters (Table 2.4). The use of default denoising parameters only identified four bacterial phyla in the fecal bacterial community of newborn beef calves. In contrast, optimization of denoising parameters identified 10 bacterial phyla in the fecal bacterial community of newborn beef calves. Phylum Proteobacteria (0.92±0.002%) dominated the fecal bacterial community when using default denoising parameters, while phylum Firmicutes (0.51±0.11%) was most abundant in the bacterial profiles obtained through optimized denoising parameters. The relative abundance of phylum Proteobacteria (0.92±0.002%) decreased ($p < 0.05$) by four times (0.22±0.06%) when using optimized denoising parameters. At the genus level, *Sphingopyxis* (0.19±0.05%), Unclassified-genera (0.19±0.07%), and *Methylobacterium-Methylorurum* (0.08±0.02%) accounted for almost 50% of bacterial genera under default denoising parameters. However, *Clostridium sensu stricto 1* (0.28±0.04%), *Escherichia-Shigella* (0.11±0.03%), and *Corynebacterium* (0.01±0.02%) represent around 50% of bacterial genera when using optimized denoising parameters. Optimization of denoising parameters increased the total number of taxa identified in the fecal bacterial community from 103 to 199 taxa (Table 2.4). Optimization also increased the number of genera identified in individual animals (Table 2.5).

2.5 Discussions

Profiling of host-associated microbiota using NGS tools has gained much attention due to the vital roles they play in animal health. NGS-based high-throughput data needs to be analyzed using powerful bioinformatics pipelines. Recently, there is an increase in the field of bioinformatics to design user-friendly pipelines to cater high demand in analyzing NGS-based microbial data. QIIME2 is one the widely used open-source software to handle high-throughput microbial data (Bolyen et al., 2019; Mohsen et al., 2019). User-friendly plugins and interfaces in QIIME2 provide an opportunity for researchers in any field to use this platform to perform a microbial analysis without having prior experience (Mohsen et al., 2019). However, the outcomes of microbial community composition depend on these plugins. For example, Reeder and Knight (2010) revealed that a denoising step could reduce the diversity of microbial communities at the OTU level compared to that from a non-denoised (filtered data instead of denoising) pipeline. However, the authors reported a microbial community-dependent effect of denoising on richness. For example, the highest OTU richness was observed in the gut without denoising step, whereas, the skin microbial community had the highest richness after denoising (Reeder and Knight, 2010). This suggests that denoising can refine the microbial community assessments and remove the noises generated through sequencing. Similar to Reeder and Knight (2010), the present study revealed that optimized denoising parameters refined outcomes of microbial analysis.

In the present study, we optimized denoising parameters to ensure that forward and reverse reads have a minimum of 20 bp overlap by choosing a median Phred score value ≥25 to truncate raw reads. The recommended number for overlapped base pairs between forward and reverse reads during the read merge step should be at least 20 bp (Hall and Beiko, 2018). Generally, truncation length for forward and reverse reads has decided based on the interactive sequence quality plot where the mean Phred score begins to drop due to low-quality sequenced reads (Callahan et al., 2016; Hall and Beiko, 2018; Callahan et al., 2019; Estaki et al., 2020). In the present study, the use of a mean Phred score ≥25 to select truncation length revealed no overlap and lower merged reads when compared to optimized denoising. Optimized denoising parameters created a 29 bp overlap between forward and reverse reads. The present study revealed that the increase in merged read counts due to optimized denoising parameters provide a sufficient number of reads to conduct downstream analysis.

A previous study conducted by Hall and Beiko (2018), provides general guidelines for conducting the 16S rRNA data analysis using the QIIME2 bioinformatics pipeline using the microbial profiles generated from the gut of bumblebees. Therefore, this pipeline may not be applicable to all microbial data analyses and require customization depends on the experimental setup. For example, Rai and colleagues (2021) conducted a study with preclinical gingival samples from mice (a low microbial biomass sample) and reported that their limitations were low sample size and low-quality scores when profiling microbiota. The main challenge of the present study was to maintain an overlapped region between forward and reverse reads. We achieved overlapped region by defining quality of reads as median Phred score. The quality of sequenced reads in low microbial biomass samples is affected by background noises (Kennedy et al., 2023). For example, a study conducted by Kennedy and colleagues (2023) compared recent studies that characterized microbial populations in human fetuses and reported that all detected microbial taxa are contaminations during sample processing, DNA extraction, and sequencing (Kennedy et al., 2023). We used rectal swabs to profile the bacterial community in newborn beef calves, which are low microbial biomass sample. As a result, we observed a decrease in base quality towards the end of the reverse read compared to the forward read. Choose of median Phred score over mean compensated the base quality of reverse reads and also created an overlapped region that increased merge reads.

In the present study, optimization of denoising parameters not only improved denoising statistics and feature count but also improved the downstream analyses including, diversity (alpha and beta) and taxonomic classification. Increase in the number of samples included in the analysis following the optimization of denoising directly influenced diversity and compositional analysis. A previous study conducted by Rai and colleagues (2021) reported that the use of two different truncation lengths for microbial data analysis resulted in distinct microbial diversities when using the same data set. Another study reported that the use of appropriate quality trimming leads to an increase in the number of merged reads passed through the quality control step when using both QIIME1 and QIIME2 bioinformatics platforms (Mohsen et al., 2019). For example, increasing the quality-trimming threshold from zero to eight increased the number of reads after merging (Mohsen et al., 2019). The findings of the present study further confirmed the importance of optimization of denoising parameters and bioinformatics pipelines in microbial data analysis.

2.6 Conclusions

Denoising parameters in QIIME2 open-source software package determine the merge read counts and the number of samples that can be used in downstream analysis. Our study further confirmed that optimization of denoising parameters also enhances the accuracy of outcomes and interpretations of host-associated bacterial community compositions compared to default denoising parameters, especially when the biological sample contains low microbial biomass. These findings will help researchers to identify approaches to customize user-friendly bioinformatics pipelines to increase the accuracy of their results.

2.7 References

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018 ;6(1):1-7.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852-7.

Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11(12):2639-43.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA. QIIME allows analysis of high-throughput community sequencing data. Nat methods. 2010 ;7(5):335-6.

Estaki M, Jiang L, Bokulich NA, McDonald D, Gonzalez A, Kosciolek T, et al. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. Curr Protoc Bioinformatics. 2020;70(1):e100.

Hall M, Beiko RG. 16S rRNA gene analysis with QIIME2. Microbiome analysis: methods protoc. 2018:113-29.

Kennedy KM, de Goffau MC, Perez-Munoz ME, Arrieta MC, Backhed F, Bork P, et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. Nature. 2023;613(7945):639-49.

Kogut MH, Arsenault RJ. Editorial: Gut Health: The New Paradigm in Food Animal Production. Front Vet Sci. 2016;3:71.

Li F, Li C, Chen Y, Liu J, Zhang C, Irving B, Fitzsimmons C, Plastow G, Guan LL. Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. Microbiome. 2019 ;7:1-7.

Malmuthuge N, Griebel PJ. Fetal environment and fetal intestine are sterile during the third trimester of pregnancy. Vet Immunol Immunopathol. 2018;204:59-64.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal. 2011 ;17(1):10-2.

McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217.

Mohsen A, Park J, Chen YA, Kawashima H, Mizuguchi K. Impact of quality trimming on the efficiency of reads joining and diversity analysis of Illumina paired-end reads in the context of QIIME1 and QIIME2 microbiome analysis frameworks. BMC Bioinformatics. 2019;20(1):581.

Nakandalage R, Guan L L, Malmuthuge N, et al. Microbial Interventions to Improve Neonatal Gut Health Microorganisms. 2023; 11, no. 5: 1328.

Pluske JR, Turpin DL, Kim JC. Gastrointestinal tract (gut) health in the young pig. Anim Nutr. 2018;4(2):187-96.

Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS One. 2020;15(1):e0227434.

Rai SN, Qian C, Pan J, Rai JP, Song M, Bagaitkar J, et al. Microbiome data analysis with applications to pre-clinical studies using QIIME2: Statistical considerations. Genes Dis. 2021;8(2):215-23.

Reeder J, Knight R. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Methods. 2010;7(9):668-9.

Rosen MJ, Callahan BJ, Fisher DS, Holmes SP. Denoising PCR-amplified metagenome data. BMC bioinformatics. 2012 ;13(1):1-6.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41.

Telatin A. Qiime Artifact EXtractor (Qax): A Fast and Versatile Tool to Interact with QIIME2 Archives. BioTech. 2021 ;10(1):5.

2.8 Tables and Figures

Table 2.1 Impact of optimization of denoising parameters on denoising statistics.

Parameters	Default denoising	Optimized denoising	P value
Merge read count	504 ± 104^a	$29,312\pm2,468^b$	<0.05
Percentage of merge reads (%)	1 ± 0.2^a	45 ± 2^b	<0.05
Mean ASVs	26 ± 5^a	337 ± 49^b	<0.05
Number of samples used for	D1 – 16	D1 – 20	-
downstream analysis	D2 – 01	D2 – 20	

Statistical test - one-way ANOVA and mean separation Duncan multiple range test, [a, b] means with different superscript within same column are significantly different (p<0.05). Default – forward truncation length – 281; reverse truncation length – 207, Optimized – forward truncation length – 294; reverse truncation length – 241.

Table 2.2 Merged sequence counts of individual samples under default and optimized denoising parameters.

		Sequence Count	
Animal ID	Sampling Day	Default denoising parameters	Optimized denoising parameters
ID-1	2	0	26093
ID-1	1	1268	6057
ID-2	2	26	18297
ID-2	1	520	10792
ID-3	2	59	19942
ID-3	1	0	33
ID-4	2	9	24729
ID-4	1	2196	12059
ID-5	2	282	22068
ID-5	1	2140	14276
ID-6	2	18	26972
ID-6	1	1010	8488
ID-7	2	28	19091
ID-7	1	400	11302
ID-8	2	10	18990
ID-8	1	1041	9659
ID-9	2	15	27959
ID-9	1	744	9900
ID-10	2	26	27199
ID-10	1	1307	11346
ID-11	2	6	17809

ID-11	1	270	8624
ID-12	2	7	21462
ID-12	1	455	10871
ID-13	2	101	14216
ID-13	1	642	7916
ID-14	2	8	30758
ID-14	1	2153	9839
ID-15	2	8	21525
ID-15	1	967	14398
ID-16	2	18	24173
ID-16	1	1388	9791
ID-17	2	11	24237
ID-17	1	616	12617
ID-18	2	2	30131
ID-18	1	134	8622
ID-19	2	69	21892
ID-19	1	143	17320
ID-20	2	11	23791
ID-20	1	158	5520

Table 2.3 Impact of optimization of denoising parameters on Alpha diversity indexes.

Parameters[*]	Default denoising	Optimized denoising	P value
Chao1	25±4[a]	118±9[b]	<0.0001
Shannon	3±0.3[a]	5±0.1[b]	<0.0001
Faith's phylogenetic diversity	3.5±0.1[a]	16±1.5[b]	<0.0001
Fisher's index	5.7±0.9[a]	18±1.8[b]	<0.0001
Simpson's index	0.7±0.04	0.9±0.01	NS
Pielou's evenness	0.8±0.01	0.8±0.01	NS
Good's coverage	0.99±0.01	0.99±0.01	NS

[*]alpha diversity analysis conducted for both fecal sample collection days (n=40)
Default – forward truncation length – 281; reverse truncation length – 207, Optimized – forward truncation length – 294; reverse truncation length – 241; NS-not significant.

Table 2.4 Impact of optimization of denoising parameters on taxonomic classification.

Parameters	Default denoising	Optimized denoising
Total number of identified taxa (Genus level)	103	199
Number of phyla	4	10
Predominant phyla* (%)		
Proteobacteria	(0.92±0.02)	(0.22±0.06)
Firmicutes	(0.06±0.03)	(0.51±0.12)
Actinobacteriota	(0.00±0.00)	(0.21±0.05)
Predominant genera* (%)		
Sphingopyxis	(0.19±0.05)	(0.00±0.00)
Unclassified	(0.19±0.07)	(0.00±0.00)
Methylobacterium-Methylorurum	(0.08±0.02)	(0.00±0.00)
Clostridium sensu stricto 1	(0.01±0.00)	(0.28±0.04)
Escherichia-Shigella	(0.00±0.00)	(0.11±0.03)
Corynebacterium	(0.00±0.00)	(0.01±0.02)
Number of genera identified in individual animal (ID-Animal identification number)	ID1-28	ID1-85
	ID2-23	ID2-85
	ID3-04	ID3-17
	ID4-31	ID4-115

ID5-43	ID5-116
ID6-31	ID6-88
ID7-13	ID7-56
ID8-26	ID8-85
ID9-24	ID9-70
ID10-28	ID10-97
ID11-18	ID11-82
ID12-14	ID12-64
ID13-27	ID13-84
ID14-35	ID14-87
ID15-33	ID15-90
ID16-44	ID16-89
ID17-25	ID17-68
ID18-14	ID18-75

	ID19-15	ID19-46
	ID20-14	ID20-85

Default – forward truncation length – 281; reverse truncation length – 207, Optimized – forward truncation length – 294; reverse truncation length – 241. *relative abundance
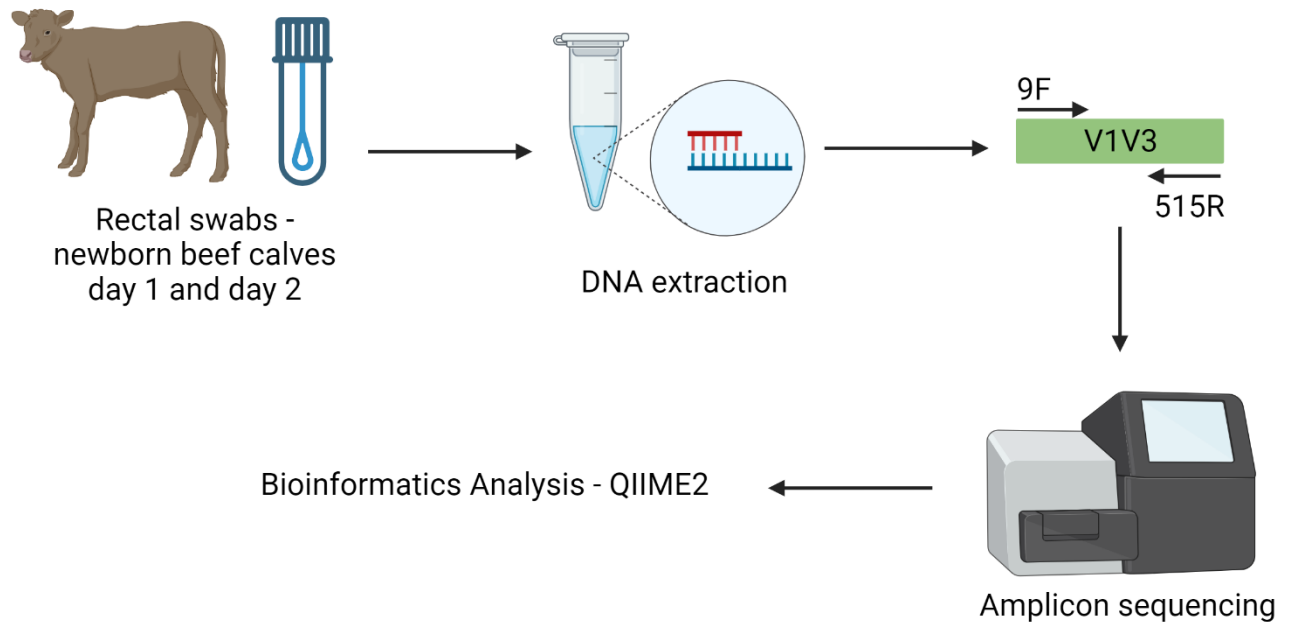
Figure 2.1 Experimental design and laboratory analysis for profiling fecal bacterial community in newborn beef calves (n = 20). Fecal samples were collected withing 24 hours after birth (Day 1 – D1) and after allow calves to suckle colostrum from dams (Day 2 – D2). Extracted DNA from rectal swabs were sequenced using V1V3 (9F – 515R) primers and data analysis were conducted using QIIME2 bioinformatics tool.
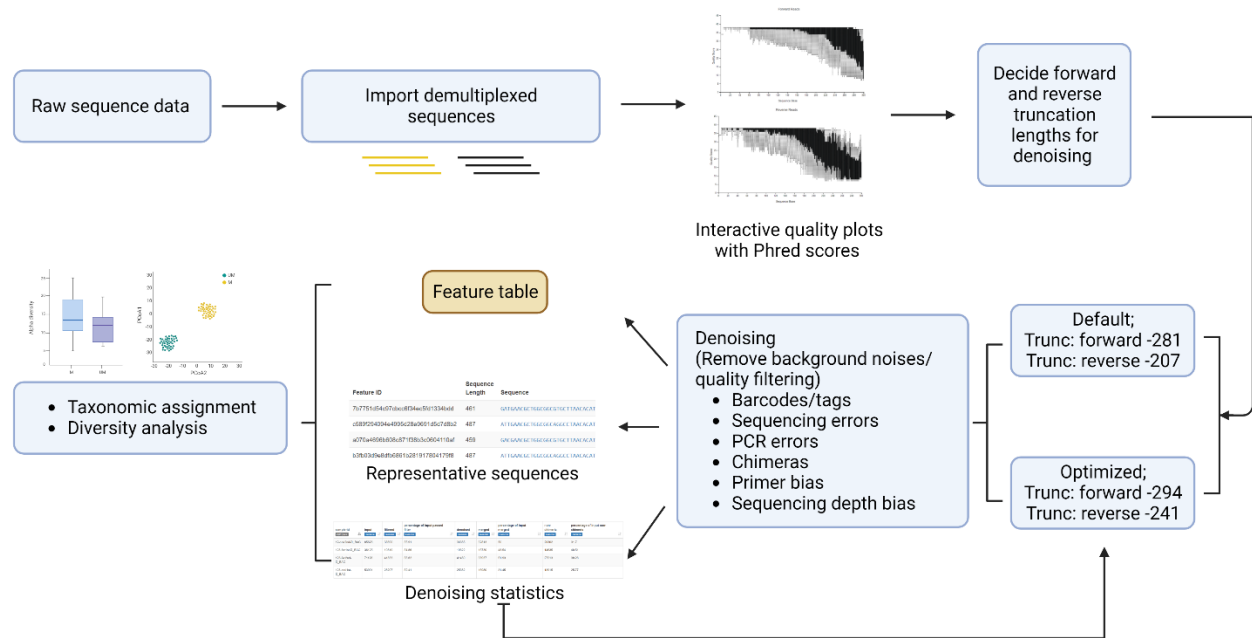
Figure 2.2 The simplified flow chart illustrate the sequence bacterial data analysis pipeline in QIIME2 bioinformatics platform. Demultiplex sequence reads were imported into QIIME2 and interactive quality plots were visualized to select the denoising parameters and quality of the sequencing reads. First, default denoising parameters (truncation: forward – 281; reverse – 207, mean Phred score - ≥25) were used for bacterial data analysis. After reviewing denoising statistics, denoising parameters were optimized (truncation: forward – 294; reverse – 241, median Phred score - ≥25) to maintain minimum overlap base pairs between forward and reverse sequence reads using DADA2 denoising plugin in QIIME2. Denoising/quality filtering allow to removal of noisy reads and merged forward and reverse reads while generating three visualizable files (feature table, representative sequence and denoising statistics. Denoising statistic file displayed the merged read counts after denoising. Finally, diversity and community analysis were performed to identify fecal bacterial community in newborn beef calves.
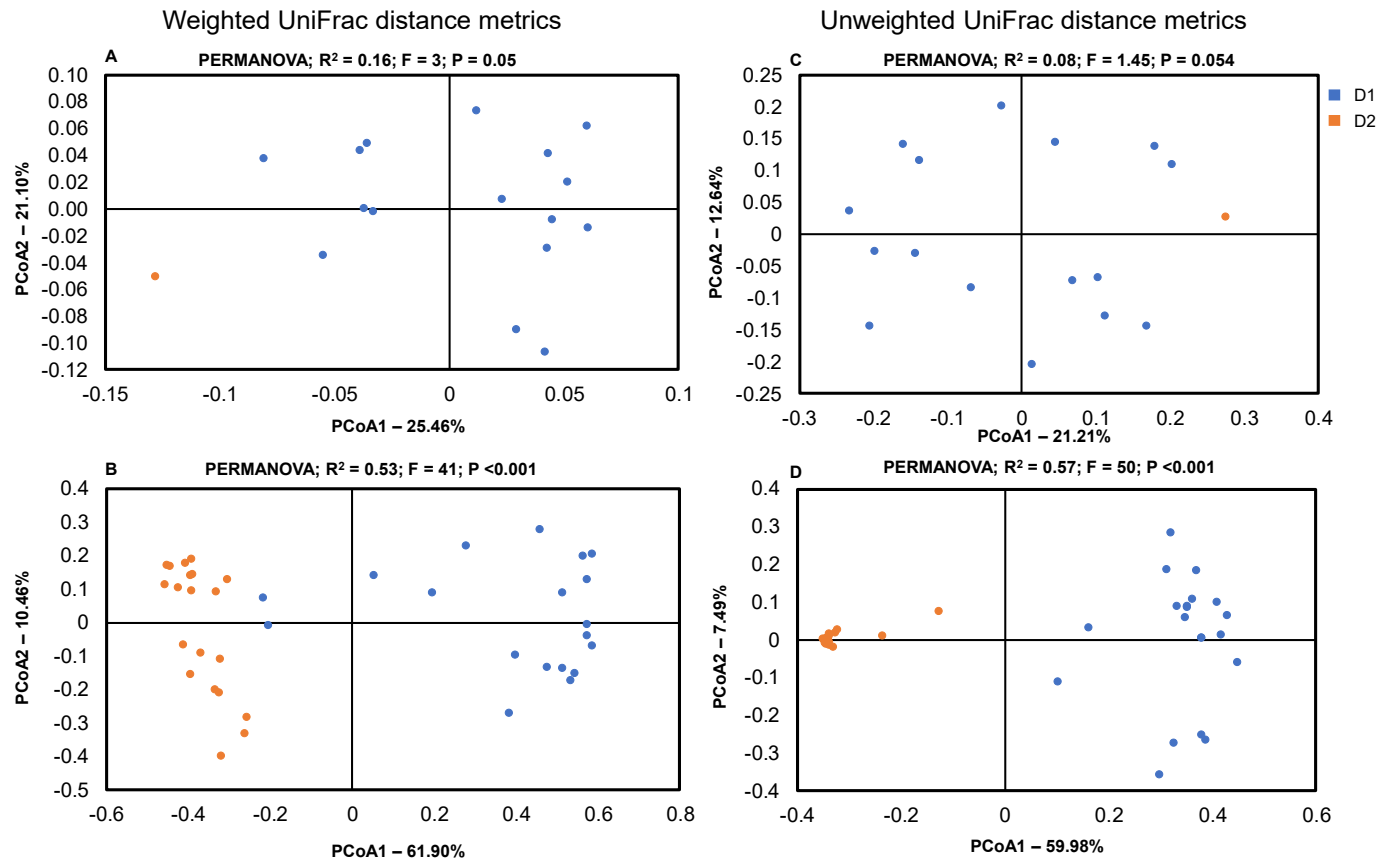
Figure 2.3 Impact of optimization of denoising parameters (default: forward – 281; reverse – 207, optimized: forward – 294; reverse – 241) on fecal bacterial community composition in day 1 (D1) and day 2 (D2) assessed using V1V3 hypervariable region of 16S rRNA gene by weighted UniFrac distance metrics (A) default and (B) optimized denoising parameters and unweighted UniFrac distance metrics (C) default and (D) optimized denoising parameters in QIIME2 platform and PERMANOVA statistic test.

Chapter 3. Selection of suitable hypervariable region and genetic material to profile true microbial community in newborn beef calves using low microbial biomass samples

3.1 Abstract

Amplicon sequencing (16S rRNA gene sequencing) is widely used to profile host-associated microbial communities. However, profiling low microbial biomass samples using amplicon sequencing is challenging due to contaminations that lead to artificial microbial sequences. This study compared amplicon sequencing-based bacterial profiles generated by different genetic materials (DNA vs. RNA) and hypervariable regions of the 16S rRNA gene (V1V3 vs. V3V4). Rectal and oral swabs (n=40) were collected from 20 newborn beef calves and used to extract DNA and RNA. Both DNA and RNA-based sequencing were performed by targeting the V3V4 region of the 16S rRNA gene. In addition, only DNA-based sequencing was performed by targeting both V1V3 and V3V4 regions of the 16S rRNA gene. All sequence runs included no template controls (NTC) and positive controls (*Clostridium butyricum*). Data were analyzed using the QIIME2 platform with customized quality filtering steps to remove environmental contaminations and filter out low-abundant amplicon sequencing variants (ASVs). Sequencing analysis revealed that sequences generated from NTC could be assigned to bacterial taxa irrespective of the genetic materials and target regions, suggesting that the amplicon sequencing process can introduce contaminations. When comparing the impact of the target region, the richness and evenness were higher ($p<0.05$) in the fecal and oral bacterial profiles generated from the V1V3 region compared to those of the V3V4 region. Taxonomic assignment of bacterial profiles generated using two hypervariable regions revealed distinct bacterial communities. For example, Actinobacteria (fecal - 0.41±0.09%; oral - 0.51±0.10%) was abundant in bacterial profiles generated from the V1V3 region, whereas Firmicutes (fecal - 0.37±0.11%; oral - 0.34±0.10%) was abundant in those of V3V4 region when comparing D1 samples. When comparing different genetic materials, DNA-based bacterial profiles (both oral and fecal) had higher diversity compared to RNA-based profiles on D1. In contrast, the diversity of the RNA-based profiles was higher than DNA-based profiles on D2. This suggests the presence of a diverse active bacterial community when samples were collected on D2 when compared to D1. In conclusion, the diversity and composition of bacterial communities derived from low microbial biomass samples depend on the choice of genetic materials and

the hypervariable region of the 16S rRNA gene. The inclusion of appropriate controls is crucial to increasing the accuracy of results, regardless of the sequencing technique.

3.2 Introduction

Colonization of gut microbiota begins during the birthing process when the fetus exposes to the outside environment (Caballero-Flores et al., 2022). The early life microbial colonization of the gastrointestinal tract (GIT) in humans and livestock species gain more attention over the last decade due to its influence on overall host health (Kogut et al., 2016; Pluske et al., 2018). Early-life microbiota plays an important role in priming host immune responses (Malmuthuge and Guan, 2017; Ravisankar et al., 2018; Nakandalage et al., 2023) and microbial dysbiosis in neonates has been linked to negative health outcomes. As a result, studies to develop potential interventions to manipulate microbial composition have been increased (Raman et al., 2019; Kim et al., 2021; Rosa et al., 2021; Song et al., 2021; Slanzon et al., 2022; Nuzhat et al., 2023). However, the development and implementation of early-life microbial interventions require a clear understanding of true microbial colonization in the GIT. Profiling of a true microbial community depends on various factors, including the choice of the hypervariable region of the 16S rRNA gene and genetic materials.

The choice of the hypervariable region of the 16S rRNA gene affects microbial profiling. Nine hypervariable regions in the 16S rRNA marker gene displayed sequence diversity and distinct taxonomic resolution (the ability of different regions to identify microbial taxa) among different hypervariable regions (Chakravorty et al., 2007; Abellan-Schneyder et al., 2021). The most used regions to profile the gut microbiota in literature are the V1V3 and V3V4 regions (Abellan-Schneyder et al., 2021). For example, a study comparing the gut microbial diversity between two hypervariable regions reported a higher alpha diversity in the V1V2 region compared to the V3V4 region (Kameoka et al., 2021), suggesting the identification of differences in microbial communities when using different hypervariable regions to profile microbial communities. However, there is a lack of knowledge on the most suitable hypervariable region to profile bacterial communities of newborn calf feces and mouth, when samples are derived using swabs.

Moreover, microbial community composition depends on the type of genetic material used to generate amplicons. Amplicon sequencing-based microbial profiling mainly uses DNA, which represents active, dead, and dormant microbial cells (De Vrieze et al., 2018; Salgar-Chaparro and Machuca, 2019; Wang et al., 2019). Dead and dormant cells detected by DNA-based amplicon sequencing lead to an overestimation of active microbial cells in a community (Wang et al., 2019). In contrast, RNA-based amplicon sequencing

can profile only active/alive microbes (De Vrieze et al., 2018). Thus, profiling of microbial communities using RNA-based amplicon sequencing can identify active microbial communities and will provide an opportunity to develop successful microbial interventions.

There is heterogeneity among previous microbiota studies in beef cattle due to the choice of primer pairs, genetic material used for sequencing, sequencing approaches, data analysis techniques, and other environmental factors (Li et al., 2016; Weese and Jelinski, 2016). Most of the previous microbial assessments used fecal samples, tissue samples, and digesta samples to profile microbial communities. However, swabbing is one of the non-invasive sampling techniques that can easily perform, transport, and store with less complexity compared to other sampling techniques (Reyman et al., 2019; Radhakrishnan et al., 2023). A study comparing human gut microbial composition between rectal swabs and fecal samples reported a higher correlation between the two sampling methods (Radhakrishnan et al., 2023), suggesting the potential of using swabs for sample collection. The collection of fecal samples can be challenging in newborns due to the less availability of fecal samples immediately after birth (Reyman et al., 2019). Rectal swabs can be used to collect microbial samples easily at any time point, consistently. Thus, in this study, we want to see whether swab samples can be used to profile the bacterial community of newborn beef calves. The objectives of this study were to (1) assess the impact of genetic materials (DNA vs. RNA) on microbial profiles generated from low microbial biomass samples, and (2) assess the impact of hypervariable regions of the 16S rRNA gene (V1V3 vs. V3V4) on oral and fecal microbial communities of newborn beef calves.

3.3 Materials and Methods

3.3.1   Animal experiments and sample collection

All experiment protocols were approved by the Livestock Care Committee at the University of Alberta (AUP-00004183) and the University of Saskatchewan (AUP - 20170015) and were conducted following the guidelines of the Canadian Council on Animal Care. Newborn beef calves (n=20) were obtained from the Rayner Dairy Research and Teaching Facility, University of Saskatchewan (Saskatoon, SK) within 24 hours of birth. Oral and rectal swabs of newborn beef calves were collected after birth prior to colostrum feeding (sample collected on Day 1-D1). Calves were allowed to suckle colostrum from cows and re-sample after

24 hours (sample collected on Day 2-D2). All samples were snap frozen in liquid nitrogen and stored at -80°C for nucleic acid extraction.

3.3.2 Nucleic acid extraction

Total genomic DNA from oral and rectal swabs was extracted using the QIAamp Fast DNA stool Mini kit (Qiagen, USA), with a fast spin-column procedure. Briefly, swab samples were suspended in 1 mL InhibitEx buffer and lysed protein using 25 µL of proteinase K. DNA was eluted using elution buffer after several cleaning steps following the manufacturer's instruction.

Total RNA was extracted from oral and rectal swabs using the miRNeasy mini kit (Qiagen, USA). Samples were transferred to QIAzol lysis reagent and incubated at room temperature for 5 min. After incubation with chloroform, centrifuge for 15 min at 12,000 g at 4°C, and RNA was extracted using RNeasy mini-column to RNase-free water. Afterward, total RNA was reverse transcribed to cDNA using iScript™ reverse transcription supermix (Bio-Rad, CA), following the manufacturer's instructions. The quality and quantity of the DNA and RNA were evaluated using the Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Delaware, USA).

3.3.3 Profiling of oral and fecal bacterial communities in newborn beef calves using amplicon sequencing of the 16S rRNA gene

Diluted DNA (50ng/µL) was used to amplify the V1V3 and V3V4 hypervariable regions of the 16S rRNA gene separately using universal bacterial primers (Table 3.1). Total of 30 µL of diluted cDNA (50ng/µL) was used to amplify V3V4 region (Table 3.1). Amplicons were barcoded and sequenced on an Illumina Miseq 300 with 300 bp paired-end reads at the McGill University and Genome Quebec Innovation Center (Quebec, Canada). Similar to the samples, a positive control (*Clostridium butyricum*) and a no template control (NTC; nucleus-free water) were processed in the same PCR and sequencing run at the same time.

3.3.4 Estimation of bacterial densities in DNA and RNA based amplicon sequencing using quantitative real time PCR

DNA and RNA-based quantitative real-time PCR was performed to estimate the total bacterial density using universal primers (U2-F: ACTCCTACGGGAGGCAG; U2-R: GACTACCAGGGTATCTAATCC) (Stevenson and Weimer, 2007) and SYBR green chemistry (Fast SYBR Green Master Mix; Applied Biosystems) with the ViiA 7 Real-time PCR System (Applied Biosystems, CA, USA). Cycle conditions were as follows: the

holding stage at 95°C for 5 minutes, followed by 40 cycles at 95°C for 20 seconds and 60°C for 1minute. The melting curve conditions were as follows: 95°C for 15 seconds, followed by 60°C for 1 minute, and 95°C for 15 seconds. The standard curves for total bacteria were generated using serial dilutions of purified plasmid containing the 16S rRNA gene of *Butyrivibrio hungatei*. The copy number of 16S rRNA genes of targeted bacteria per swab was calculated using the same equation mentioned in Zhou et al. (2009).

3.3.5   Bioinformatics Analysis

Demultiplexed paired-end 16S rRNA gene sequences (oral and fecal samples) were denoised (Table 3.2; at least 20bp merge; median quality score >25) using DADA2 in QIIME2. Feature tables generated after denoising were subjected to three different quality filtering steps; 1. filtering low abundant ASVs <0.005% of samples), 2. filtering environmental contamination (mitochondria, chloroplast, cyanobacteria, chloroflexi, archaea), and 3. Filtering ASVs present in NTC to remove contaminated and artificial ASVs. Then, the remaining good-quality sequences were used to assign taxonomy using the SILVA database. SILVA138-99 classifier was trained separately (using train-classifier feature in QIIME2 platform) for V1V3 and V3V4 primers prior to taxonomic assignment. Diversity analysis was performed using the diversity plugin in QIIME2 bioinformatics software. Bacterial diversity of oral and fecal microbiota was performed using alpha diversity indices (Shannon and Chao1) and beta diversity analysis (Weighted UniFrac Distance Matrix) within QIIME2 (2022.2 version).

3.3.6   Statistical Analysis

Bacterial community composition and diversity between V1V3 and V3V4 regions were compared using DNA-based bacterial profiles. DNA and RNA-based bacterial communities were compared using profiles generated by targeting V3V4 hypervariable region. Alpha diversity indices were analyzed using the non-parametric Kruskall-Wallis test and statistical significances were declared at *p<0.05*. Beta diversity analysis was performed using the weighted UniFrac distance matrices and PERMANOVA statistical test in the QIIME2 environment and statistical significances were declared at *p<0.001*.

3.4 Results

3.4.1 Comparison of fecal bacterial profiles generated by targeting varying regions of the 16S rRNA gene

3.4.1.1 Differences in read counts and observed amplicon sequence variants (ASVs) between V1V3 and V3V4 hypervariable regions

Comparison of read counts between the two primer pairs (V1V3 vs. V3V4 hypervariable regions) revealed a higher number of raw and denoised read counts ($p < 0.05$) for bacterial profiles generated from V3V4 compared to V1V3 region when samples were collected on D1 (Table 3.3). Denoised read counts were lower in bacterial profiles generated from the V3V4 region compared to the V1V3 region when samples were collected on D2 (Table 3.3). However, there was no significant effect of hypervariable regions on read counts when samples were collected on D2. In NTC, raw and denoising read counts were higher in bacterial profiles generated from the V3V4 region compared to the V1V3 region (Table 3.3). Comparison of ASVs between sequences generated from amplicon targeting two hypervariable regions showed a higher number of ASVs in bacterial profiles generated from the V1V3 region compared to the V3V4 region on both sampling days (D1 and D2). In contrast, amplicon sequencing of the V1V3 region generated lower ASVs counts in NTC compared to the V3V4 region.

3.4.1.2 Comparison of fecal microbial diversity between V1V3 and V3V4 primer pairs

The profiling of bacterial community in newborn beef calves significantly affects the choice of the hypervariable region of the 16S rRNA gene. The richness of the bacterial community (Chao 1) was higher ($p < 0.05$) in sequences generated from amplicons targeting the V1V3 region compared to the V3V4 region on both sampling days (Figure 3.2). Similarly, a higher ($p < 0.05$) evenness and richness (Shannon) of fecal bacterial diversity was observed in bacterial profiles generated from the V1V3 region when compared to V3V4 when samples were collected on D1. The evenness of the bacterial community was higher but not significant when samples were collected on D2 (Figure 3.2).

The bacterial communities generated by targeting the V1V3 region clustered apart from those generated by targeting the V3V4 region. PERMANOVA statistical test showed that 87% of the observed variation ($R^2 = 0.87$; F = 526; $p < 0.001$) between the bacterial communities was explained by the hypervariable region (Figure 3.3A). A similar pattern was observed on both sampling days (Figure 3.3A).

Similar to the PCoA plot, the phylogenetic tree displayed two clear phylogenetically distinct clusters when comparing the bacterial communities between V1V3 and V3V4 hypervariable regions (Figure 3.3B).

3.4.1.3 Taxonomic assessment of fecal bacterial community between V1V3 and V3V4 regions

Based on taxonomic classification, the V1V3 region identified ten phyla when samples were collected on D1 and four phyla when samples were collected on D2. In contrast, sequences generated from amplicon targeting the V3V4 region identified eleven and five bacterial phyla when samples were collected on D1 and D2, respectively. Actinobacteria (0.41±0.09%), followed by Firmicutes (0.22±0.08%) and Proteobacteria (0.21±0.07%) dominated bacterial profiles generated from V1V3 region when samples were collected on D1 ($p< 0.05$). In contrast, bacterial profiles generated from the V3V4 region were dominated by Firmicutes (0.37±0.11%), followed by Actinobacteria (0.24±0.05%), and Proteobacteria (0.20±0.07%) when samples were collected on D1. After 24 hours of first sampling (on D2), the V1V3 region showed a higher abundance ($p< 0.05$) of Firmicutes (0.75±0.14%) followed by Proteobacteria (0.21±0.04%), while Proteobacteria (0.49±0.04%) and Firmicutes (0.45±0.10%) were predominant when sequences generated from amplicon targeting V3V4 region. At the genus level, *Corynebacterium* dominated the fecal bacterial profiles generated by targeting the V1V3 region, whereas *Escherichia-shigella* ($p<0.05$) dominated those of the V3V4 region (Figure 3.4C). Furthermore, unclassified taxonomies were highly abundant sequences generated from amplicon targeting V3V4 compared to V1V3 hypervariable region in both sampling days (Figure 3.4C, Figure 3.5).

A total of 162 and 194 bacterial genera were identified from V1V3 and V3V4 regions, respectively when samples were collected on D1 (Figure 3.4A). After 24 hours (D2), a total of 39 and 30 bacterial genera were observed when sequences were generated from amplicon targeting V1V3 and V3V4 regions, respectively (Figure 3.4B). Moreover, V1V3 and V3V4 hypervariable regions shared a total of 115 and 22 bacterial genera when samples were collected on D1 and D2, respectively (Figure 3.4A and B). Comparison of fecal bacterial genera in individual animals also revealed that some bacterial genera were unique when sequences generated from amplicon targeting V1V3 and V3V4 regions and some bacterial genera were commonly identified in both hypervariable regions (Figure 3.4C, Table 3.7).

3.4.2   Comparison of the fecal bacterial profiles generated using different genetic materials

3.4.2.1 Differences in read counts and observed ASVs between different genetic materials

A comparison of read counts between the two different genetic materials revealed a higher number of ($p<0.05$) raw and denoised read counts in DNA-based amplicon sequencing compared to RNA-based amplicon sequencing when samples were collected on D1 (Table 3.4). In contrast, lower ($p<0.05$) raw and denoised read counts were observed in DNA-based sequencing compared to RNA-based sequencing when samples were collected on D2. In NTC, DNA-based sequencing generated higher raw and denoised read counts compared to RNA-based sequencing. A higher ($p<0.05$) number of ASVs was identified in DNA-based sequencing compared to RNA-based sequencing when samples were collected on D1 (Table 3.4), while ASV counts were not statistically different between DNA and RNA-based sequencing when samples were collected on D2. ASVs observed in NTC were almost five times lower in RNA-based sequencing compared to DNA-based sequencing (Table 3.4).

3.4.2.2 Comparison of the fecal bacterial diversity using DNA and RNA-based amplicon sequencing

There was a significant effect of genetic materials on the fecal bacterial diversity in newborn beef calves (Figure 3.6). The bacterial richness and evenness were higher ($p<0.05$) when using DNA-based sequencing to profile bacteria compared to RNA-based sequencing when samples were collected on D1 (Figure 3.6 A and C). In contrast, bacterial richness and evenness were lower ($p<0.05$) when using DNA-based sequencing to profile bacteria compared to RNA-based sequencing when samples were collected on D2 (Figure 3.6 B and D).

PERMANOVA statistical test revealed that only 5.4% of the observed variation between bacterial communities could be explained by the genetic material (R^2 = 0.054; $p<0.002$; F = 4.3; Figure 3.7A). The bacterial profiles between DNA and RNA-based sequencing cluster together without having clear separation. These findings were further confirmed by observing no clear phylogenetic clusters between DNA and RNA-based profiles (Figure 3.7B).

3.4.2.3 Taxonomic assessment of the fecal bacterial community between DNA and RNA-based amplicon sequencing

Based on taxonomic classification, DNA-based sequencing identified eight phyla, while RNA-based sequencing captured five phyla when samples were collected on D1. Five and four bacteria phyla were

identified from DNA and RNA-based sequencing, respectively when samples were collected on D2. Among the identified phyla, Firmicutes (0.34±0.11%) was predominant in DNA-based amplicon sequencing, followed by Actinobacteria (0.17±0.04%), and Proteobacteria (0.13±0.06%) when samples were collected on D1. In contrast, a higher relative abundance of Proteobacteria (0.33±0.08%) was observed in RNA-based sequencing when samples were collected on D1. Proteobacteria dominated both DNA (0.44±0.04%) and RNA (0.32±0.08%) based sequencing followed by Firmicutes (DNA-0.28±0.05%; RNA-0.20±0.08%) when samples were collected on D2.

Unclassified genera were identified in both DNA-based (0.29±0.06%) profiling and RNA-based (0.70±0.26%) profiling (Figure 3.8C; *p*< 0.05). However, the relative abundance of unclassified genera was two times higher in RNA-based (0.70±0.26%) profiling compared to DNA-based profiling when samples were collected on D1 (Figure 3.8C, Figure 3.9). The most abundant genus in DNA-based profiling was *Escherichia-shigella* (0.43±0.03%), while unclassified (0.59±0.16%) genera were predominant in RNA-based profiling when samples were collected on D2. A total of 71 and 197 bacterial genera were identified in RNA and DNA-based profiling, respectively when samples were collected on D1 (Figure 3.8A). After 24 hours (on D2), a total of 30 bacterial genera were identified from DNA-based profiling, while RNA-based profiling generated 24 bacterial genera (Figure 3.8B). Even though 45 bacterial genera were shared between DNA and RNA-based communities when samples were collected on D1, only 13 bacterial genera were shared when samples were collected on D2. Comparison of fecal bacterial genera in individual animals also revealed that some bacterial genera were unique to DNA and RNA and some bacterial genera commonly identified by both genetic materials (Figure 3.8C, Table 3.7).

3.4.2.4 Comparison of DNA and RNA based bacterial densities

Estimation of the fecal bacterial densities in newborn beef calves revealed a lower (*p*<0.05) density when using RNA-based profiling compared to DNA-based profiling irrespective of the sample collection day (Figure 3.10). RNA-based bacterial densities were lower by 10,000 and 1,000 times on D1 and D2, respectively compared to DNA-based fecal bacterial densities in newborn beef calves.

3.4.3 Comparison of the oral bacterial profiles generated by targeting different regions of the 16S rRNA gene

3.4.3.1 Differences in read counts and observed amplicon sequence variants (ASVs) between V1V3 and V3V4 regions

Comparison of read counts between two hypervariable regions (V1V3 vs. V3V4) identified a higher ($p<0.05$) number of read counts in the V3V4 region compared to the V1V3 region irrespective of the sampling time point (Table 3.5). The V1V3 region identified a higher number of ASVs ($p<0.05$) compared to the V3V4 region when samples were collected on D1 and D2 (Table 3.5). In NTC, the number of ASVs generated from sequences generated from amplicon targeting V3V4 region was almost double that of the sequences generated from amplicon targeting V1V3 region (Table 3.5).

3.4.3.2 Comparison of the oral bacterial diversity between V1V3 and V3V4 primer pairs

A higher evenness and richness were observed in the oral bacterial profiles generated by targeting the sequences generated from amplicons targeting the V1V3 region compared to the sequences generated from amplicons targeting the V3V4 region in both sampling time points (Figure 3.11). However, there was no significant effect of hypervariable region on oral bacterial diversity of newborn beef calves (Figure 3.10 B, C, and D), except for bacterial richness when samples were collected on D1 ($p<0.05$; Figure 3.11A).

The oral bacterial communities captured in sequences generated from amplicon targeting V1V3 and V3V4 regions showed two different clusters in the PCoA plot (Figure 3.12A). PCoA plots represent almost 95% of the data (Figure 3.12A). PERMANOVA statistical test revealed that 84% ($R^2 = 0.84$; F = 389; $p<0.001$) of the observed variation between the oral bacterial communities could be explained by the hypervariable region (Figure 3.12A). Similarly, the phylogenetic tree also displayed two phylogenetically different clusters between sequences generated from amplicons targeting V1V3 and V3V4 regions for the oral bacterial community (figure 3.12B).

3.4.3.3 Taxonomic assessment of oral bacterial community in V1V3 and V3V4 regions

Based on the taxonomic assessment, the oral bacterial community in newborn beef calves consisted of 12 and 11 bacterial phyla in sequences generated from amplicons targeting V1V3 and V3V4 regions, respectively when samples were collected on D1. The sequences generated from amplicons targeting the V1V3 region identified 10 bacterial phyla and the sequences generated from amplicons targeting the V3V4

region captured 12 bacterial phyla when samples were collected on D2. When samples were collected On D1, three bacterial phyla accounted for 88% of the oral bacteria communities generated by the sequences generated from amplicons targeting the V1V3 region (Actinobacteriota – 0.51±0.10%; Proteobacteria – 0.20±0.06%; Firmicutes – 0.17±0.06%; $p<$ 0.05). When the sequences generated from amplicons targeting the V3V4 region were used to profile the oral bacteria, Firmicutes (0.34±0.10%), followed by Actinobacteriota (0.28±0.05%), and Proteobacteria (0.19±0.05%) accounted for the majority of bacterial phyla when samples were collected on D1. When samples were collected on D2, Proteobacteria dominated ($p<$ 0.05) the oral bacterial communities generated by sequencing both V1V3 (0.54±0.12%) and V3V4 (0.34±0.04%) regions. When samples were collected on D2, the second most abundant phyla in the bacterial profiles generated by sequencing the V1V3 region were Bacteroidetes (0.21±0.09%). Whereas, Firmicutes (0.27±0.09%) was the second most abundant phyla (0.19±0.06%) in the bacterial profiles generated by sequencing the V3V4 region.

At the genus level, *Corynebacterium* dominated the bacterial profiles generated by sequencing V1V3 (0.29±0.04%) and V3V4 (0.13±0.04%) regions when sample collected on D1 (Figure 3.12C). When sample collected on D2 *Mannheimia* (0.18±0.02%), *Porphyromonas* (0.13±0.03%), *Alysiella* (0.13±0.03%) and *Neisseria* (0.12±0.02%) represented around 50% of represented around 50% of oral bacterial genera in the bacterial profiles generated by sequencing V1V3 region. However, *Bibersteinia* (0.34±0.03%) and *Streptococcus* (0.17±0.03%) represented almost 50% of the oral bacterial profiles generated by sequencing the V3V4 region when sample collected on D1. In addition, a higher relative abundance of unclassified genera was observed in the bacterial profiles generated by sequencing the V3V4 region compared to bacterial profiles generated by sequencing the V1V3 region on both sampling days (Figure 3.13C, Figure 3.14).

When samples were collected on D1, a total of 168 and 185 oral bacterial genera were identified from the bacterial profiles generated by sequencing V1V3 and V3V4 regions, respectively (Figure 3.13A). When samples were collected on D2, the V1V3 region captured 98 bacterial genera while the V3V4 region identified 137 bacterial genera (Figure 3.13B). Two hypervariable regions shared 120 and 72 oral bacterial genera when samples were collected on D1 and D2, respectively. Comparison of oral bacterial genera in

individual animals also revealed that some bacterial genera were unique to V1V3 and V3V4 regions and some bacterial genera commonly identified by both hypervariable regions (Figure 3.13C, Table 3.8).

3.4.4 Comparison of the oral bacterial communities using different genetic materials

3.4.4.1 Differences in read counts and observed ASVs between DNA and RNA-based sequencing

Similar to the fecal bacterial communities, the use of DNA-based sequencing to profile the oral bacteria generated a higher ($p<0.05$) number of raw and denoised reads compared to RNA-based profiling when samples were collected on D1. When samples were collected on D2, a higher ($p<0.05$) number of raw and denoised read counts were observed in RNA-based profiling compared to DNA-based profiling (Table 3.6). Moreover, a higher number of ASV counts was observed in DNA-based profiling compared to RNA-based profiling when samples were collected on D1 ($p<0.05$) and D2 ($p>0.05$). In NTC, ASV counts were six times higher in DNA-based profiling compared to RNA-based profiling.

3.4.4.2 Comparison of oral bacterial diversity between DNA and RNA-based sequencing

There was a significant effect of genetic materials on the oral bacterial diversity of newborn beef calves (Figure 3.15). The richness (Chao1) of the oral bacterial community was higher ($p<0.05$) in DNA-based profiling compared to RNA-based profiling when samples were collected on D1 (Figure 3.15A). However, the richness of the oral bacterial community was similar between DNA and RNA-based profiling when samples were collected on D2 (Figure 3.15B). The evenness and richness of the oral bacterial community were higher ($p<0.05$) in DNA-based profiling compared to RNA-based profiling in both sampling points (Figure 3.15C, D).

PCoA plots represent around 45% of the data (Figure 3.16A). PERMANOVA statistical test revealed that only 16% (R^2 = 0.16; F = 15; $p<0.001$) of the variation between the bacterial communities was explained by the genetic material used for bacterial profiling (Figure 3.16A). The bacterial communities identified using DNA and RNA-based profiling cluster together (Figure 3.16A). There were no clear phylogenetically distinct clusters identified from the oral bacterial communities generated with different genetic materials (Figure 3.16B).

3.4.4.3 Taxonomic assessment of oral bacterial community between DNA and RNA-based sequencing

The oral bacterial community of newborn beef calves on D1 consisted of 11 and 15 bacterial phyla when profiled using DNA and RNA, respectively. When samples were collected on D2, 12 bacterial phyla were identified in both DNA and RNA-based profiling. Three bacterial phyla accounted for nearly 80% of the oral bacteria identified using DNA-based profiling when samples were collected on D1 (Firmicutes – 0.34±0.07%; Actinobacteria – 0.28±0.05%; Proteobacteria – 0.19±0.06%). More than 50% of the oral bacterial community profiled using RNA on D1 belong to Proteobacteria (0.54±0.14%; $p< 0.05$), followed by Bacteroidetes (0.22±0.07%) and Firmicutes (0.12±0.05%). When samples were collected on D2, Proteobacteria (0.53±0.10%) and Bacteroidetes (0.48±0.15%) dominated the DNA and RNA-based profiling, respectively.

At the genus level, *Corynebacterium* (0.25±0.09%) dominated the DNA-based bacterial profiles, while *Psychrobacter* (0.27±0.05%) and *Pseudomonas* (0.13±0.02%) represented 40% of bacterial genera in RNA-based profiles when samples were collected on D1 (Figure 3.14C). *Bibersteinia* (0.34±0.03%) and *Streptococcus* (0.17±0.02%) accounted ($p< 0.05$) for 51% of bacterial genera on D2 when DNA was used as the genetic material for profile oral bacteria. *Porphyromonas* (0.33±0.06%), *Bibersteinia* (0.12±0.04%), and *Neisseria* (0.11±0.03%) represented 56% of bacterial genera in the oral bacterial community on D2 when RNA was used as the genetic material for profile oral bacteria. Even though both DNA and RNA-based profiling identified unclassified genera, DNA-based profiling showed a higher relative abundance of unclassified genera compared to RNA-based oral bacterial profiling (Figure 3.17C, Figure 3.18).

When samples were collected on D1, a total of 185 and 154 bacterial genera were identified in DNA and RNA-based profiling, respectively (Figure 3.17A). When samples were collected on D2, DNA-based profiling generated 137, while RNA-based profiling identified 99 bacterial genera (Figure 3.17B). DNA and RNA-based profiling shared 103 and 68 bacterial genera when samples were collected on D1 and D2, respectively. Comparison of oral bacterial genera in individual animals also revealed that some bacterial genera were unique to DNA and RNA and some bacterial genera commonly identified by both genetic materials (Figure 3.17C, Table 3.8).

3.4.4.4 Comparison of DNA and RNA based oral bacterial densities

Estimation of the oral bacterial densities in newborn beef calves revealed a lower ($p<0.05$) density when using RNA compared to DNA at both sampling points (Figure 3.19). RNA-based bacterial densities were 1,000 and 10,000 times lower when samples were collected on D1 and D2, respectively compared to DNA-based oral bacterial densities in newborn beef calves.

3.5 Discussions

The present study revealed that the use of oral and rectal swabs to profile bacterial communities of newborn beef calves can contain contaminant taxa irrespective of the hypervariable regions (V1V3/V3V4) and genetic material used (DNA/RNA) and these contaminations can be eliminated by using appropriate controls and quality filtering step during data analysis. Swabs collect small amounts of biological samples (small amounts of genetic materials) that contain lower amounts of bacterial genetic materials. In the presence of low microbial DNA, NGS sequencing generates higher levels of artificial sequences (Kennedy et al., 2023). Contaminations that originate during microbial profiling outweigh the true microbial signals generated in low microbial biomass samples compared to high microbial biomass samples (Kennedy et al., 2023). In our study, NTCs were used during PCR and sequencing along with the fecal and oral samples to identify contaminations. Following quality filtering to remove low frequency and environmental contaminants, NTCs contained sequences that could be assigned to bacterial taxa. Previous studies reported that the use of low or no microbial biomass samples generated microbial profiles that lead to misinterpretation of host-associated microbial community composition (Eisenhofer et al., 2018; Liu et al., 2020; Heida et al., 2021; Kennedy et al., 2023). Our study observed a lower number of contaminated ASVs when using the V1V3 region to profile bacterial communities compared to the V3V4 region. Moreover, the use of RNA as the genetic material decreased the number of contaminant taxa in both fecal and oral bacterial communities. These findings suggest that researchers need to pay close attention to the choice of primers and genetic materials when profiling bacterial communities using low microbial biomass samples.

The present study further highlighted that bacterial community assessments can also be influenced by the choice of hypervariable regions of the 16S rRNA gene. The use of the V1V3 region in sequencing captured a highly diverse oral and fecal bacterial community compared to that of the V3V4 region. A

previous study also reported that targeting the V1V2 region generated highly diverse bacterial profiles from human fecal samples compared to that of the V3V4 region (Kameoka et al., 2021). Similarly, previous studies in humans have reported that V1 to V4 regions provide a higher resolution and more distinct fecal microbial profiles compared to other regions of the 16S rRNA gene (Kim et al., 2011; Chen et al., 2019). A study conducted in humans reported that the V1V3 region was more appropriate to study the dynamics of oral microbiota than the V3V4 region (Zheng et al., 2015). In agreement with the past studies, our findings also suggest that the use of the V1V3 region captures the diversity of bacterial communities accurately, even when using low microbial biomass samples.

In the present study, beta diversity analysis revealed distinct bacterial clusters between V1V3 and V3V4 hypervariable regions regardless of the sampling location. The use of phylogenetic analysis revealed these two variable regions profile phylogenetically distinct bacteria. A previous study reported that the use of different primer pairs (V1-V2, V1-V3, V3-V4, V4, V4-V5, V6-V8, and V7-V9) to profile human fecal stool samples observe primer-specific clustering of bacterial communities at the genus level (Abellan-Schneyder et al., 2021). In agreement with the previous findings, our findings also suggest that the use of different regions of the 16S rRNA gene generates phylogenetically different oral and fecal bacterial clusters. Besides, our findings suggest that profiling bacterial communities using the V1V3 region can increase accuracy and taxonomic resolution compared to the V3V4 region.

Due to the divergence and distinct resolution of hypervariable regions of the 16S rRNA gene, taxonomic classification also varies among different primer pairs. In the present study, a clear difference in taxonomic classification was observed between bacterial communities generated by profiling V1V3 and V3V4 regions. Kameoka and colleagues (2021) reported a higher relative abundance of *Akkermansia* when using the V3V4 region to profile microbiota compared to V1V2. Abellan-Schneyder and colleagues (2021) detected *Verrucomicrobia* in human stools when using the V3V4 region but not the V1V3 region. A study in newborn beef calves reported that Proteobacteria and *Pasteurellaceae* were the most abundant phylum and family, respectively in the oral microbial community when profiling bacteria using V4 hypervariable region (Barden et al., 2020). Even though this study used negative controls to identify contaminations, the authors did not filter out the ASVs identified in negative controls due to the lower read counts (Barden et al., 2020). The use of the V4 region limits the taxonomic resolution of bacterial community due to the short amplicon length.

The present study used longer amplicon lengths by combining hypervariable regions (V3V4 and V1V3), which increases our ability to obtain a higher resolution during taxonomic classification. Although Proteobacteria was one of the abundant phyla detected in the present study, the use of the V3V4 region (long amplicon size) revealed that Firmicutes and Actinobacteria were more abundant in the oral community of neonatal calves.

The present study also revealed that bacterial community profiling can be affected by the choice of genetic materials. The present study revealed that the use of DNA to profile oral and fecal bacterial communities captured higher microbial diversity compared to RNA-based profiling, except for fecal bacterial diversity when samples were collected on D2, suggesting newborn beef calves were colonized by a diverse active microbial community on D2 compared to D1. We also speculated that the higher alpha diversity in DNA-based profiling when fecal samples were collected on D1 might be a result of transient dead bacteria on calves. In addition, we speculate that part of the active microbial community might be lost during RNA reverse-transcribed into cDNA. A previous study conducted by Li and colleagues (2016) reported controversial findings. They compared the rumen microbial community composition of beef steers using DNA and RNA-based profiling and reported a higher alpha diversity in RNA-based profiling, suggesting that rumen is colonized by a diverse active bacterial community in adult cattle. However, this study (Li et al. 2016) was conducted using high microbial biomass samples (rumen samples) from beef steers. In our study, we collected low microbial biomass samples from newborn beef calves, used to optimize denoising parameters, and conducted additional quality filtering to remove contaminations. These might be reasons for identifying controversial findings between these two studies.

In the present study, beta diversity and phylogenetic analysis revealed that both DNA and RNA-based bacterial community profiles clustered together and the bacterial communities did not differ by genetic materials. A previous study also reported that microbial communities of digestate organic waste profiles using DNA and RNA cluster together based on a weighted UniFrac distance matrix (De Vrieze et al., 2018). In agreement with the previous findings, our findings suggest that microbial community profiles using DNA and RNA cluster together, even when using low microbial biomass samples. Based on current bacterial diversity analysis, we can interpret that RNA-based profiling provides the opportunity to identify active bacterial communities in oral and fecal samples of newborn beef calves.

In addition to bacterial diversity, distinct oral and fecal bacterial taxa were observed between DNA and RNA-based profiles. A previous study reported that some rumen bacterial taxa (e.g., *Desulfovibrionaceae, Elusimicrobiaceae* and *Sphaerochaetaceae*) in beef steers could only be identified using RNA-based profiling (Li et al., 2016). Our study also identified that some bacterial taxa were only observed in RNA-based microbial profiling. The current findings suggest that profiling oral and fecal bacterial communities using RNA will help to identify an active bacterial community in low microbial biomass samples with lower levels of contaminations. However, RNA-based bacterial profiles identified higher unclassified taxa in the fecal bacterial community in newborn beef calves, suggesting active bacterial communities in newborn beef calves have not been well studied and classified. Thus, future research needs to identify and classify active bacterial community.

3.6 Conclusions

Our study suggests that low microbial biomass samples (swabs) can be used to study host-associated bacterial communities of neonates. However, the choice of hypervariable region and genetic materials can affect the outcomes of bacterial profiling. In conclusion, the V1V3 hypervariable region is more suitable to profile oral and fecal bacterial communities in newborn beef calves, as it generates minimum contaminant ASVs while capturing a highly diverse bacterial community. Furthermore, the use of RNA-based amplicon sequencing can be used to identify active bacterial communities in low microbial biomass samples, with lower levels of contamination. However, RNA-based bacterial profile still lacks the ability to study fecal bacterial community composition due to the presence of higher unclassified taxa compared to DNA-based bacterial profiling, suggesting that lack of current knowledge regarding the active bacterial classification. RNA-based bacterial profiles can be used to study fecal and oral bacterial diversity and oral bacterial community composition.

3.7 References

Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, et al. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. Msphere. 2021;6(1):e01202-20.

Barden M, Richards-Rios P, Ganda E, Lenzi L, Eccles R, Neary J, et al. Maternal influences on oral and faecal microbiota maturation in neonatal calves in beef and dairy production systems. Anim Microbiome. 2020;2(1):31.

Caballero-Flores G, Pickard JM, Nunez G. Microbiota-mediated colonization resistance: mechanisms and regulation. Nat Rev Microbiol. 2022.

Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. microbiol methods. 2007;69(2):330-9.

Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MC, Wong MC, Ng SC, Chan FK, Chan PK. Impact of preservation method and 16S rRNA hypervariable region on gut microbiota profiling. Msystems. 2019 Feb 26;4(1):e00271-18.

De Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, et al. Human placenta has no microbiome but can contain potential pathogens. Nature. 2019;572(7769):329-34.

De Vrieze J, Pinto AJ, Sloan WT, Ijaz UZ. The active microbial community more accurately reflects the anaerobic digestion process: 16S rRNA (gene) sequencing as a predictive tool. Microbiome. 2018;6(1):63.

Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol. 2019;27(2):105-17.

Guzman CE, Wood JL, Egidi E, White-Monsant AC, Semenec L, Grommen SVH, et al. A pioneer calf foetus microbiome. Sci Rep. 2020;10(1):17712.

Heida FH, Kooi EMW, Wagner J, Nguyen TY, Hulscher JBF, van Zoonen A, et al. Weight shapes the intestinal microbiome in preterm infants: results of a prospective observational study. BMC Microbiol. 2021;21(1):219.

Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME j. 2011 ;5(10):1571-9.

Kameoka S, Motooka D, Watanabe S, Kubo R, Jung N, Midorikawa Y, et al. Benchmark of 16S rRNA gene amplicon sequencing using Japanese gut microbiome data from the V1-V2 and V3-V4 primer sets. BMC Genomics. 2021;22(1):527.

Kennedy KM, de Goffau MC, Perez-Munoz ME, Arrieta MC, Backhed F, Bork P, et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. Nature. 2023;613(7945):639-49.

Kennedy KM, Gerlach MJ, Adam T, Heimesaat MM, Rossi L, Surette MG, et al. Fetal meconium does not have a detectable microbiota before birth. Nat Microbiol. 2021;6(7):865-73.

Kim M, Morrison M, Yu Z. Evaluation of different partial 16S rRNA gene sequence regions for
      phylogenetic analysis of microbiomes. J Microbiol Methods. 2011;84(1):81-7.

Kogut MH, Arsenault RJ. Editorial: Gut Health: The New Paradigm in Food Animal Production. Front Vet
      Sci. 2016;3:71.

Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, et al. Comparison of placenta
      samples with contamination controls does not provide evidence for a distinct placenta microbiota.
      Microbiome. 2016;4(1):29.

Li F, Henderson G, Sun X, Cox F, Janssen PH, Guan le L. Taxonomic Assessment of Rumen Microbiota
      Using Total RNA and Targeted Amplicon Sequencing Approaches. Front Microbiol. 2016;7:987.

Li F, Li C, Chen Y, Liu J, Zhang C, Irving B, Fitzsimmons C, Plastow G, Guan LL. Host genetics influence
      the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle.
      Microbiome. 2019 ;7:1-7.

Lim ES, Rodriguez C, Holtz LR. Amniotic fluid from healthy term pregnancies does not harbor a
      detectable microbial community. Microbiome. 2018;6(1):87.

Liu Y, Li X, Zhu B, Zhao H, Ai Q, Tong Y, et al. Midtrimester amniotic fluid from healthy pregnancies has
      no microorganisms using multiple methods of microbiologic inquiry. Am J Obstet Gynecol.
      2020;223(2):248 e1- e21.

Malmuthuge N, Griebel PJ. Fetal environment and fetal intestine are sterile during the third trimester of
      pregnancy. Vet Immunol Immunopathol. 2018;204:59-64.

Malmuthuge N, Guan LL. Understanding the gut microbiome of dairy calves: Opportunities to improve
      early-life gut health. J Dairy Sci. 2017;100(7):5996-6005.

Mishra A, Lai GC, Yao LJ, Aung TT, Shental N, Rotter-Maskowitz A, et al. Microbial exposure during early
      human development primes fetal immune cells. Cell. 2021;184(13):3394-409 e20.

Nakandalage R, Guan L L, Malmuthuge N, et al. Microbial Interventions to Improve Neonatal Gut
      Health Microorganisms. 2023*;* 11, no. 5: 1328.

Nuzhat S, Hasan SMT, Palit P, Islam MR, Mahfuz M, Islam MM, et al. Effects of probiotic and synbiotic
      supplementation on ponderal and linear growth in severely malnourished young infants in a
      randomized clinical trial. Sci Rep. 2023;13(1):1845.

Pluske JR, Turpin DL, Kim JC. Gastrointestinal tract (gut) health in the young pig. Anim Nutr.
      2018;4(2):187-96.

Rackaityte E, Halkias J, Fukui EM, Mendoza VF, Hayzelden C, Crawford ED, et al. Viable bacterial
      colonization is highly limited in the human intestine in utero. Nat Med. 2020;26(4):599-607.

Radhakrishnan ST, Gallagher KI, Mullish BH, Serrano-Contreras JI, Alexander JL, Miguens Blanco J,
      Danckert NP, Valdivia-Garcia M, Hopkins BJ, Ghai A, Ayub A. Rectal swabs as a viable alternative
      to faecal sampling for the analysis of gut microbiota functionality and composition. Sci. Rep. 2023
      ;13(1):493.

Raman AS, Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Subramanian S, et al. A sparse covarying unit that describes healthy and impaired human gut microbiota development. Science. 2019;365(6449).

Ravisankar S, Tatum R, Garg PM, Herco M, Shekhawat PS, Chen YH. Necrotizing enterocolitis leads to disruption of tight junctions and increase in gut permeability in a mouse model. BMC Pediatr. 2018;18(1):372.

Reyman M, van Houten MA, van Baarle D, Bosch AA, Man WH, Chu ML, Arp K, Watson RL, Sanders EA, Fuentes S, Bogaert D. Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. Nat. commun. 2019 ;10(1):4997.

Rosa F, Michelotti TC, St-Pierre B, Trevisi E, Osorio JS. Early life fecal microbiota transplantation in neonatal dairy calves promotes growth performance and alleviates inflammation and oxidative stress during weaning. Animals. 2021 ;11(9):2704.

Salgar-Chaparro SJ, Machuca LL. Complementary DNA/RNA-Based Profiling: Characterization of Corrosive Microbial Communities and Their Functional Profiles in an Oil Production Facility. Front Microbiol. 2019;10:2587.

Slanzon GS, Ridenhour BJ, Parrish LM, Trombetta SC, Moore DA, Sischo WM, McConnel CS. Effects of a farm-specific fecal microbial transplant (FMT) product on clinical outcomes and fecal microbiome composition in preweaned dairy calves. Plos one. 2022 ;17(10):e0276638.

Song SJ, Wang J, Martino C, Jiang L, Thompson WK, Shenhav L, et al. Naturalization of the microbiota developmental trajectory of Cesarean-born neonates after vaginal seeding. Med (N Y). 2021;2(8):951-64 e5.

Stevenson DM, Weimer PJ. Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Appl Microbiol Biotechnol. 2007;75(1):165-74.

Stinson L, Hallingstrom M, Barman M, Viklund F, Keelan J, Kacerovsky M, et al. Comparison of Bacterial DNA Profiles in Mid-Trimester Amniotic Fluid Samples From Preterm and Term Deliveries. Front Microbiol. 2020;11:415.

Wang J, Tian S, Yu H, Wang J, Zhu W. Response of Colonic Mucosa-Associated Microbiota Composition, Mucosal Immune Homeostasis, and Barrier Function to Early Life Galactooligosaccharides Intervention in Suckling Piglets. J Agric Food Chem. 2019;67(2):578-88.

Weese JS, Jelinski M. Assessment of the Fecal Microbiota in Beef Calves. J Vet Intern Med. 2017;31(1):176-85.

Zheng W, Tsompana M, Ruscitto A, Sharma A, Genco R, Sun Y, et al. An accurate and efficient experimental approach for characterization of the complex oral microbiota. Microbiome. 2015;3:48.

Zhou MI, Hernandez-Sanabria E, Guan LL. Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. AEM. 2009 ;75(20):6524-33.

3.8 Tables and Figures

Table 3.1 Bacterial primers used to profile oral and fecal bacterial communities of newborn beef calves.

Hypervariable region	Primer name	Product size (bp)	Illumina adapter sequence	Primers (5'-3')	Reference
V1V3	BAC9F	506	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT	GAGTTTGATCMTG GCTCAG	Li et al., 2019
	BAC515R		GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT	CCGCGGCKGCTGG CAC	
V3V4	341F	464	ACACTCTTTCCCTACACGAC GCTCTTCCGATCT	CCTACGGGNGGC WGCAG	Herlemann et al., 2011
	805R		GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT	GACTACHVGGGTA TCTAATCC	

Table 3.2 Denoising parameters to identify oral and fecal bacterial communities in newborn beef calves.

Samples*	Parameters	V1V3	V3V4	
		DNA	DNA	RNA
Fecal	Forward -truncation	294	271	264
	Reverse -truncation	241	222	222
Oral	Forward -truncation	291	278	269
	Reverse -truncation	246	223	225

*median Phred quality score ≥25

Table 3.3 Comparison of read counts and ASVs before and after quality filtering in V1V3 and V3V4 hypervariable regions of rectal swabs (mean±SE).

Parameters	V1V3				V3V4			
	D1	D2	NTC	PC	D1	D2	NTC	PC
Raw read count	45,101±3,099[a]	80,349±2,666	12,437	72,551	78,864±2,036[b]	86,887±2,310	43,521	95,481
Denoised read count	24,348±1,664[a]	48,266±1,811	7,122	43,838	46,454±1,595[b]	46,395±1,288	18,629	37,739
ASVs before filtered	335±76	340±64[a]	87	13	226±47	190±43[b]	146	16
ASVs feature filtered[*]	121±16	121±11[a]	31	11	85±14	72±12[b]	117	16
ASVs taxonomy filtered[**]	120±16	120±11[a]	29	11	84±14	71±12[b]	115	16
ASVs NTC filtered	116±15	116±10[a]	-	7	83±14	70±12[b]	-	14

[*] total frequency <0.005%; [**] taxonomy related to environment – mitochondria, chloroplast, cyanobacteria, chloroflexi, archaea; D1- day 1; D2- day 2; NTC - no template control; PC - positive control (*Clostridium butyricum*); Statistical test- one-way ANOVA and mean separation Duncan multiple range test, [a, b] means with different superscript within same column are significantly different ($p<0.05$).

Table 3.4 Comparison of read counts and ASVs before and after quality filtering using DNA and RNA-based bacterial profiling in V3V4 hypervariable region using rectal swabs in newborn beef calves (mean±SE).

Parameters	DNA				RNA			
	D1	D2	NTC	PC	D1	D2	NTC	PC
Raw read count	78,864±2,036[a]	86,887±2,310[a]	43,521	95,481	44,941±4,635[b]	114,733±9,081[b]	17,811	134,983
Denoised read count	46,454±1,595[a]	46,395±1,288[a]	18,629	37,739	17,817±3,026[b]	84,310±7,976[b]	8,388	112,698
ASVs before filtered	226±47[a]	190±43	146	16	90±18[b]	109±23	23	22
ASVs feature filtered*	85±14[a]	72±12	117	16	50±7[b]	59±6	19	22
ASVs taxonomy filtered**	84±14[a]	71±12	115	16	50±7[b]	59±6	19	22
ASVs NTC filtered	83±14[a]	70±12	-	14	49±7[b]	58±6	-	21

* total frequency <0.005%; ** taxonomy related to environment – mitochondria, chloroplast, cyanobacteria, chloroflexi, archaea; D1- day 1; D2- day 2; NTC - no template control; PC - positive control (*Clostridium butyricum*); Statistical test- one-way ANOVA and mean separation Duncan multiple range test, [a, b] means with different superscript within same column are significantly different ($p<0.05$).

Table 3.5 Read count and amplicon sequence variants (ASVs) before and after quality filtering in V1V3 and V3V4 hypervariable region in oral swabs.

Parameters	V1V3				V3V4			
	D1	D2	NTC	PC	D1	D2	NTC	PC
Raw read count	49,991±7,326[a]	44,562±6,048[a]	12,437	72,551	80,939±2,620[b]	83,852±1,857[b]	43,521	95,481
Denoised read count	25,321±3,897[a]	25,507±3,500[a]	6,879	42,770	45,412±1,645[b]	43,677±1,539[b]	17,187	36,204
ASVs before filtered	395±65[a]	116±23	86	16	240±18[b]	99±18	146	12
ASVs feature filtered	189±24[a]	74±10	41	15	123±7[b]	63±8	118	12
ASVs taxonomy filtered	187±24[a]	73±10	39	15	121±7[b]	63±8	117	12
ASVs NTC filtered	184±23[a]	71±10	-	11	121±6[b]	62±8	-	11

[*] total frequency <0.005%; [**] taxonomy related to environment – mitochondria, chloroplast, cyanobacteria, chloroflexi, archaea; D1- day 1; D2- day 2; NTC - no template control; PC - positive control (*Clostridium butyricum*); Statistical test- one-way ANOVA and mean separation Duncan multiple range test, [a, b] means with different superscript within same column are significantly different ($p<0.05$).

Table 3.6 Comparison of read counts and ASVs before and after quality filtering using DNA and RNA -based profiling in V3V4 hypervariable region of oral swabs (mean±SE).

Parameters	DNA				RNA			
	D1	D2	NTC	PC	D1	D2	NTC	PC
Raw read count	80,939±2,620[a]	83,852±1,857[a]	43,521	95,481	67,460±4,080[b]	98,371±5,163[b]	17,811	134,983
Denoised read count	45,412±1,645[a]	43,677±1,539[a]	17,187	36,204	37,318±3,275[b]	63,865±4,892[b]	8,297	110,332
ASVs before filtered	240±18	99±18	146	12	244±21	113±21	24	24
ASVs feature filtered	123±7[a]	63±8	118	12	78±5[b]	50±4	18	24
ASVs taxonomy filtered	121±7[a]	63±8	117	12	77±5[b]	50±4	18	23
ASVs NTC filtered	121±6[a]	62±8	-	11	76±5[b]	49±4	-	20

[*] total frequency <0.005%; [**] taxonomy related to environment – mitochondria, chloroplast, cyanobacteria, chloroflexi, archaea; D1- day 1; D2- day 2; NTC - no template control; PC - positive control (*Clostridium butyricum*); Statistical test- one-way ANOVA and mean separation Duncan multiple range test, [a, b] means with different superscript within same column are significantly different ($p<0.05$).

Table 3.7 Common and unique taxa (genus level) in fecal bacterial community profiled by different
hypervariable regions and genetic materials for individual animals.

	Number of common genera		Number of unique genera			
Animal ID	DNA & RNA	V1V3 & V3V4	DNA	RNA	V1V3	V3V4
ID1	10	42	62	5	43	42
ID2	14	50	50	14	35	23
ID3	10	14	28	17	3	24
ID4	18	49	65	17	66	48
ID5	5	40	78	5	76	53
ID6	11	40	68	11	48	53
ID7	11	19	57	5	37	45
ID8	19	44	59	9	41	45
ID9	7	29	57	2	41	44
ID10	10	36	63	5	61	50
ID11	7	41	76	3	41	55
ID12	9	18	31	12	46	32
ID13	14	21	28	8	63	25
ID14	14	31	56	9	56	51
ID15	13	28	38	11	62	31
ID16	7	34	57	5	55	38
ID17	20	28	49	7	40	51
ID18	6	30	59	3	45	42
ID19	10	15	29	13	31	27
ID20	8	30	50	1	55	36

Table 3.8 Common and unique taxa (genus level) in oral bacterial community profiled by different hypervariable regions and genetic materials for individual animals.

Animal ID	Number of common genera		Number of unique genera			
	DNA & RNA	V1V3 & V3V4	DNA	RNA	V1V3	V3V4
ID1	24	37	46	26	33	33
ID2	21	37	47	16	31	18
ID3	4	1	2	44	5	38
ID4	35	57	57	36	35	57
ID5	40	52	63	17	0	10
ID6	11	11	1	32	1	70
ID7	11	15	8	25	4	66
ID8	15	17	7	51	5	61
ID9	31	49	57	27	39	26
ID10	26	41	54	15	39	26
ID11	37	50	44	33	31	30
ID12	35	30	37	30	42	36
ID13	18	21	9	37	6	39
ID14	25	50	58	2	33	26
ID15	34	59	65	20	40	17
ID16	32	47	52	21	37	20
ID17	36	52	66	38	50	30
ID18	30	41	50	30	39	15
ID19	33	41	40	20	32	15
ID20	29	48	71	12	52	27
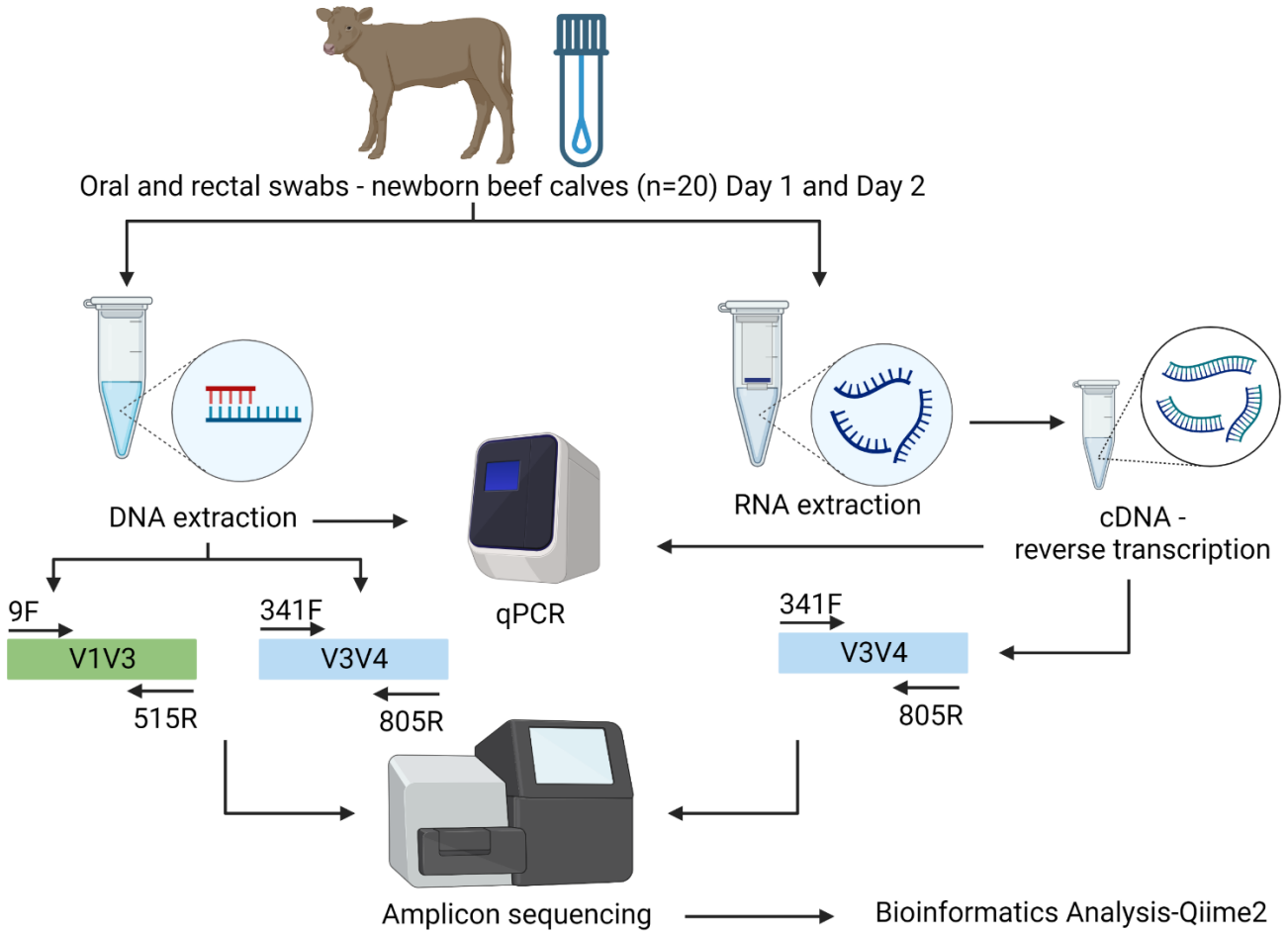
Figure 3.1 Experimental design and laboratory analysis for profiling bacterial community. Oral and fecal samples were collected from newborn beef calves (n=20) before (D1-day 1) and after (D2-day 2) colostrum feeding. Oral and rectal swabs were used to extract the genetic material for (DNA – V1V3 and V3V4 primers and RNA – V3V4 primers). qPCR was performed to estimate total bacterial densities. Amplicon sequencing (Illumina Miseq – 2×300 bp) was performed to profile bacterial community.
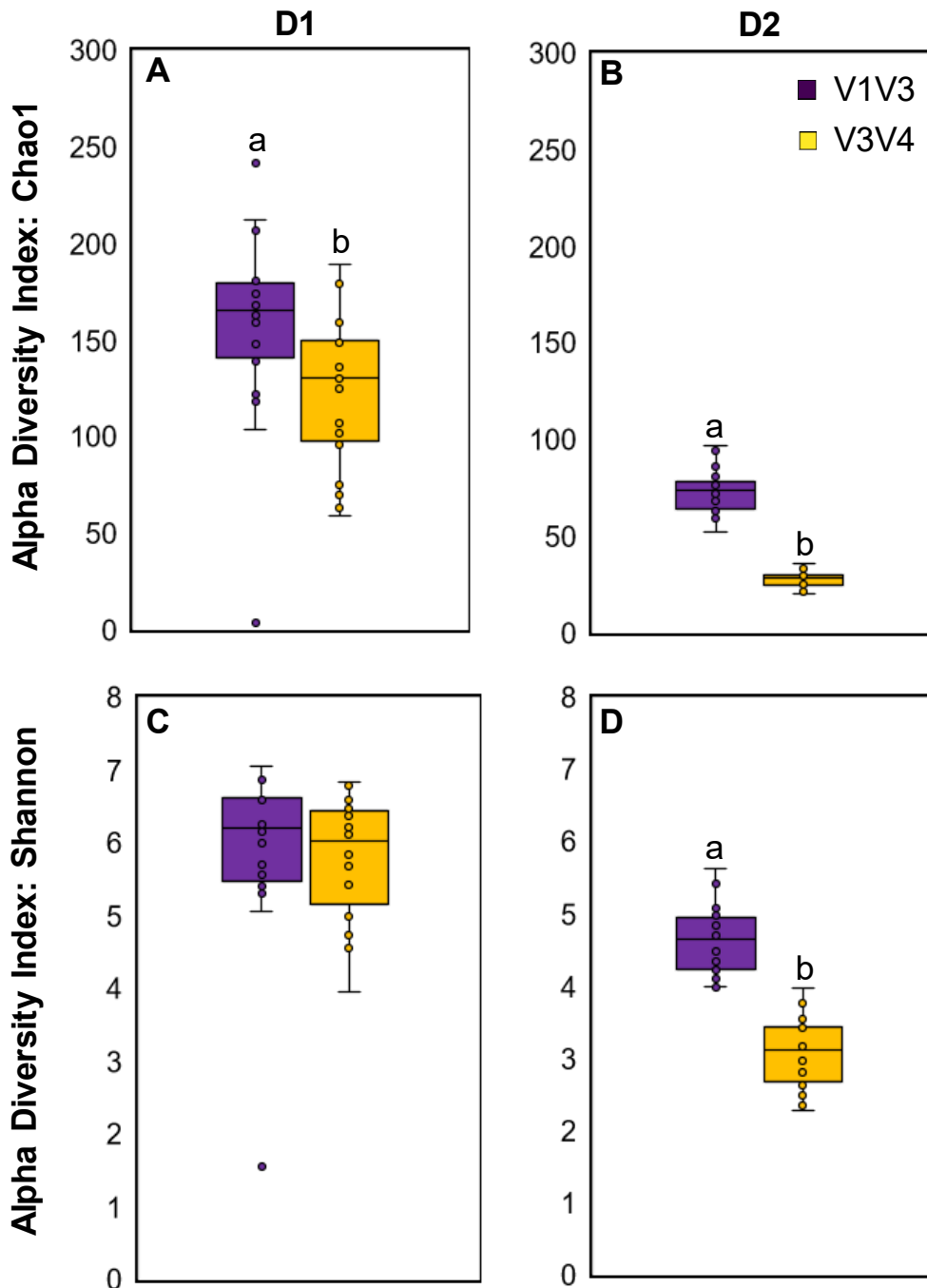
Figure 3.2 Fecal bacterial diversity comparison (DNA-based amplicon sequencing) between V1V3 and V3V4 hypervariable regions using (A) alpha diversity index Chao 1 on day one (D1), (B) on day two (D2), (C) alpha diversity index Shannon on day one and (D) on day two; [a, b] means with different superscript are significantly different at $p<0.05$.
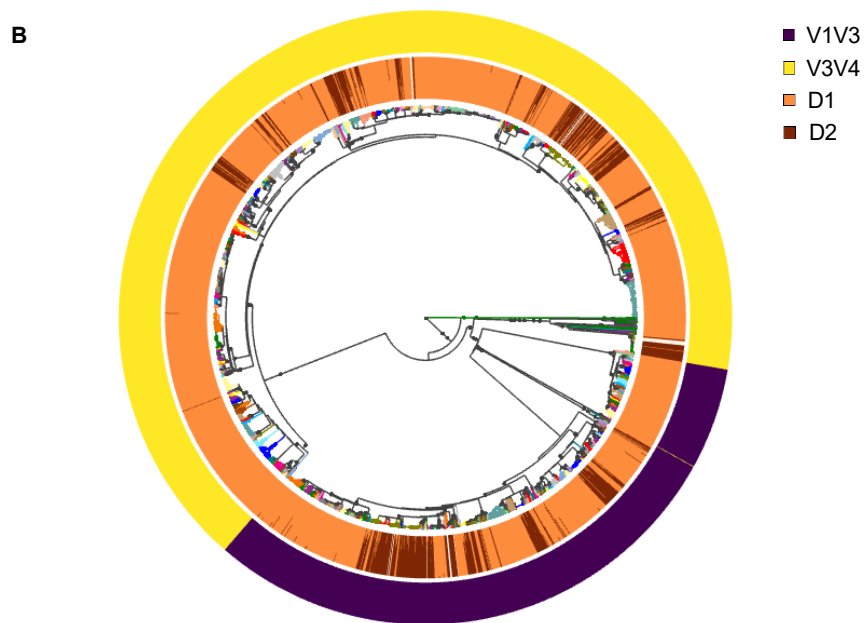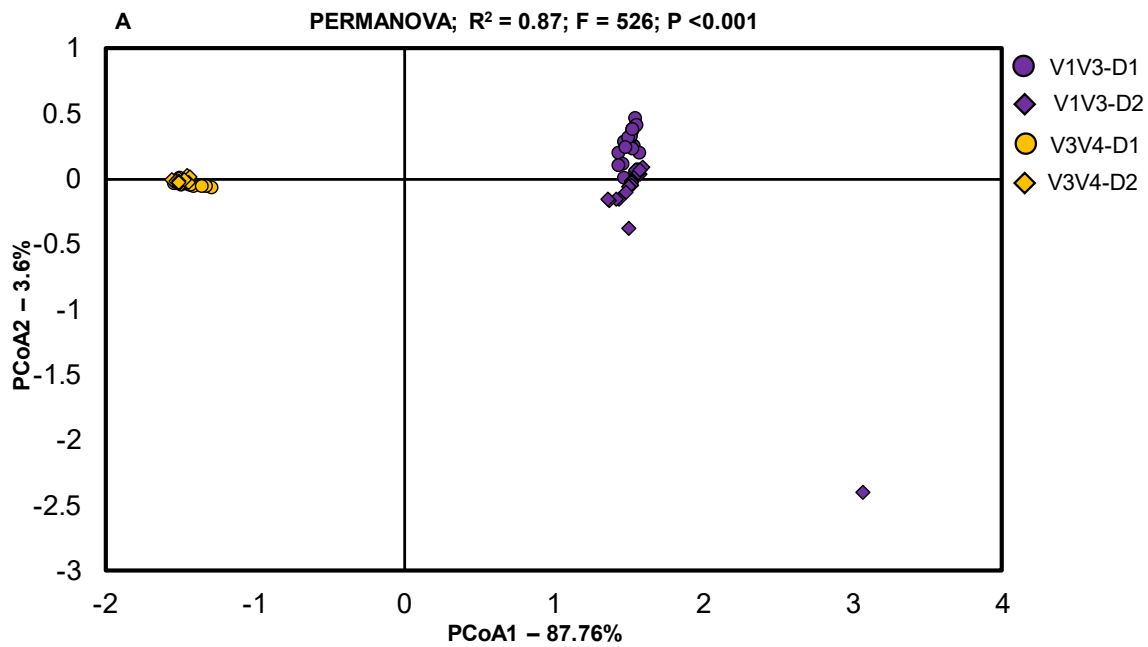
Figure 3.3 Comparison of fecal bacterial communities generated by targeting V1V3 and V3V4 hypervariable regions (A) principle coordinate analysis (PCoA) using weighted UniFrac distance metrics in QIIME2 platform and PERMANOVA statistic test and (B) phylogenetic tree shows the phylogenetic similarity and distances of bacterial taxa identified from different hypervariable regions. The outer most bar plot represents different hypervariable regions and the inner bar plot represents sampling timepoints.
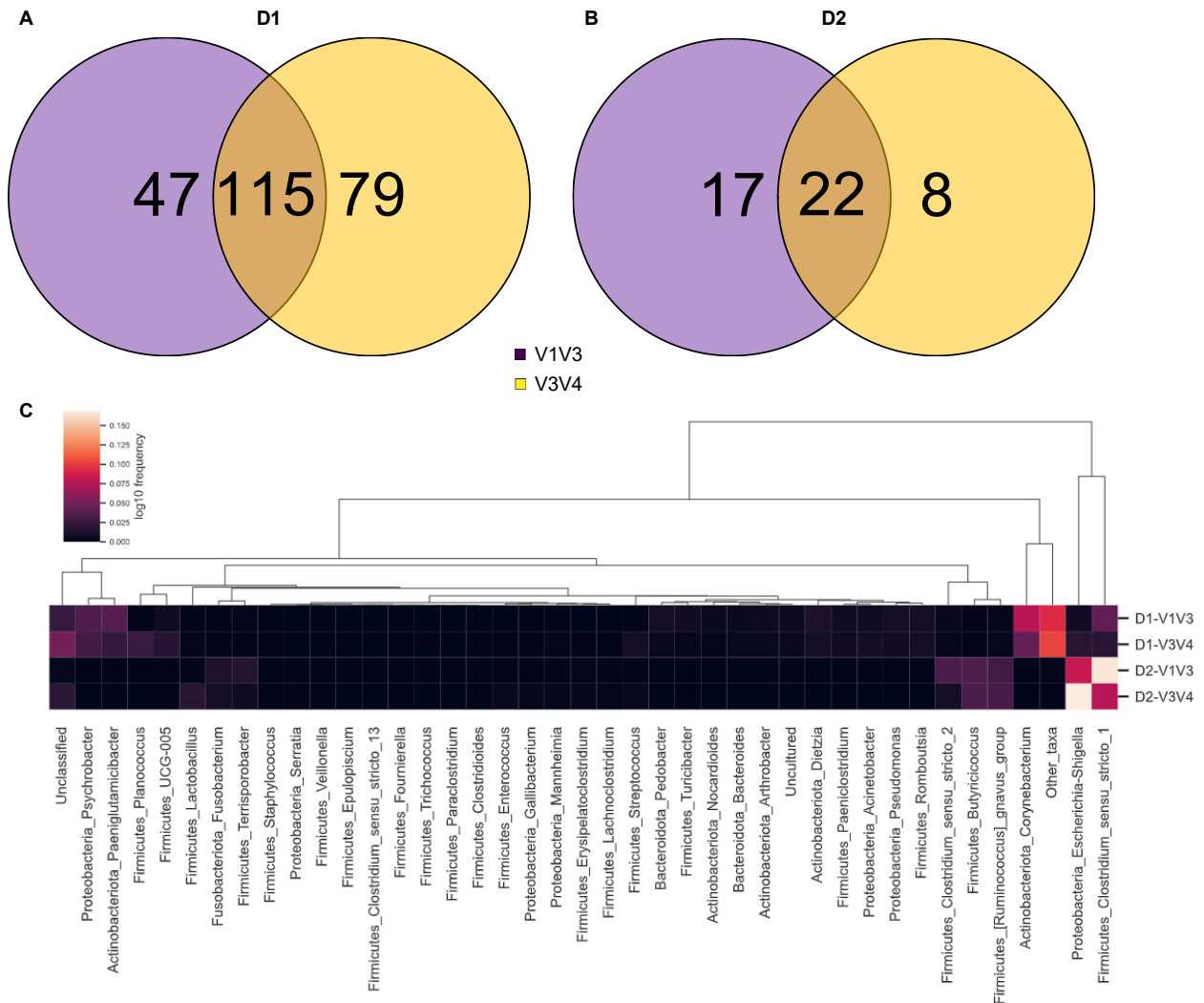
Figure 3.4 Taxonomic classification of the fecal bacterial communities generated by targeting two hypervariable regions (V1V3 and V3V4) of the 16S rRNA marker gene. (A) Venn diagrams represent the total number of unique and shared genera present in all calves (n=20) on D1, (B) on D2 and (C) heat map compare the relative abundance of forty (40) bacterial genera identified using two primer pairs. Bacterial taxa are arranged from the highest relative abundance to lowest (phylum_genus).
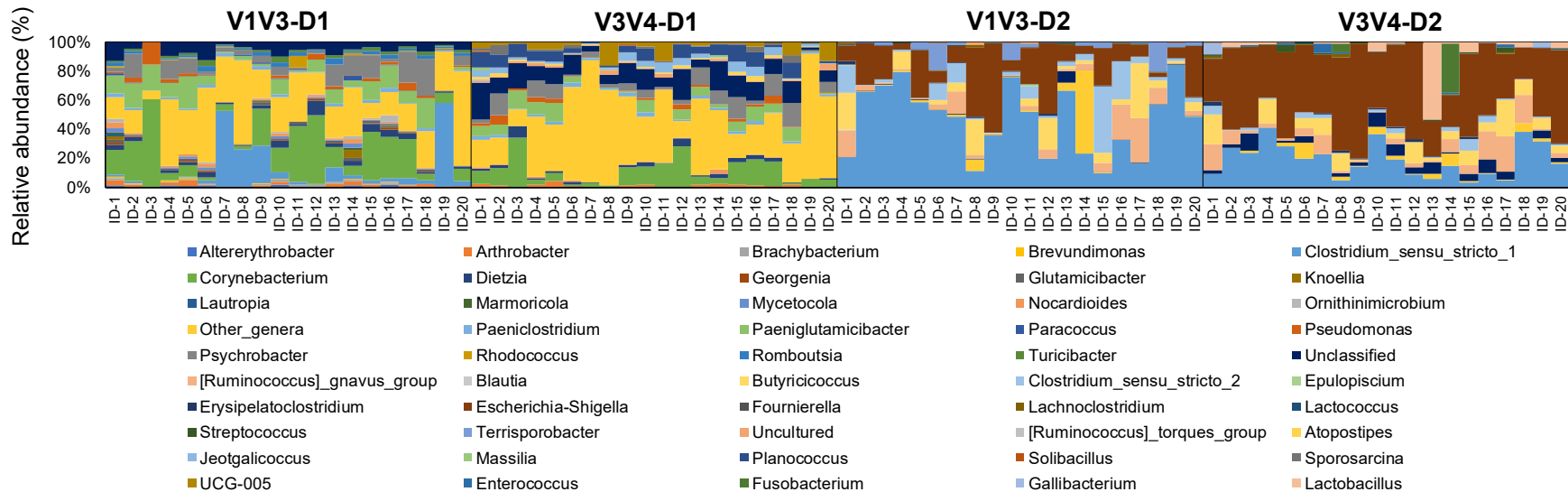
Figure 3.5 Relative abundance (for individual animals; n = 20) of the fecal bacterial communities generated by targeting two hypervariable regions (V1V3 and V3V4) of the 16S rRNA marker gene (genus level).
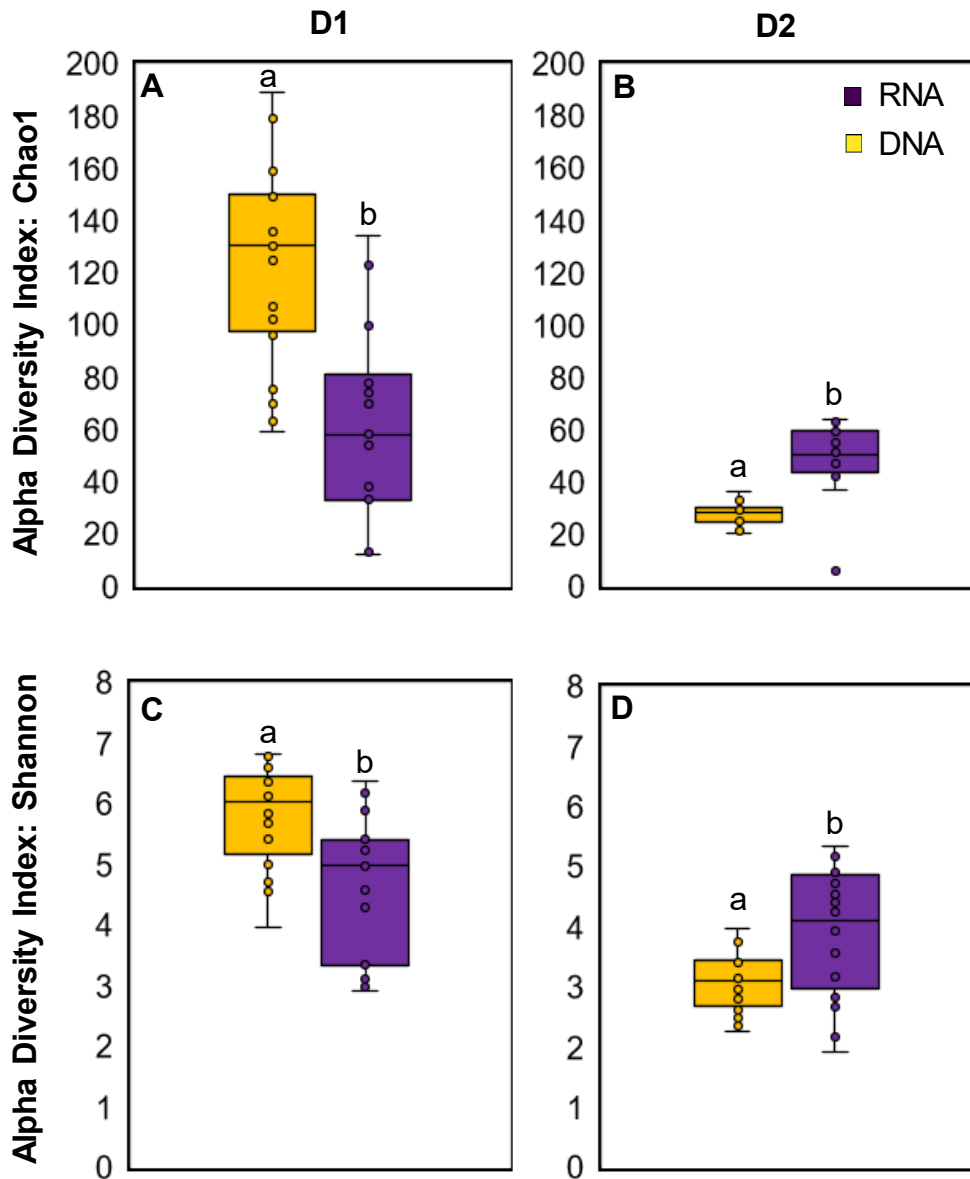
Figure 3.6 Fecal bacterial diversity comparison between DNA and RNA-based amplicon sequencing of 16S rRNA marker gene targeting V3V4 region (A) alpha diversity index Chao 1 on day 1 (D1), (B) on day 2 (D2), (C) alpha diversity index Shannon on day one (D1) and (D) on day two (D2); [a, b] means with different superscript are significantly different at $p<0.05$.
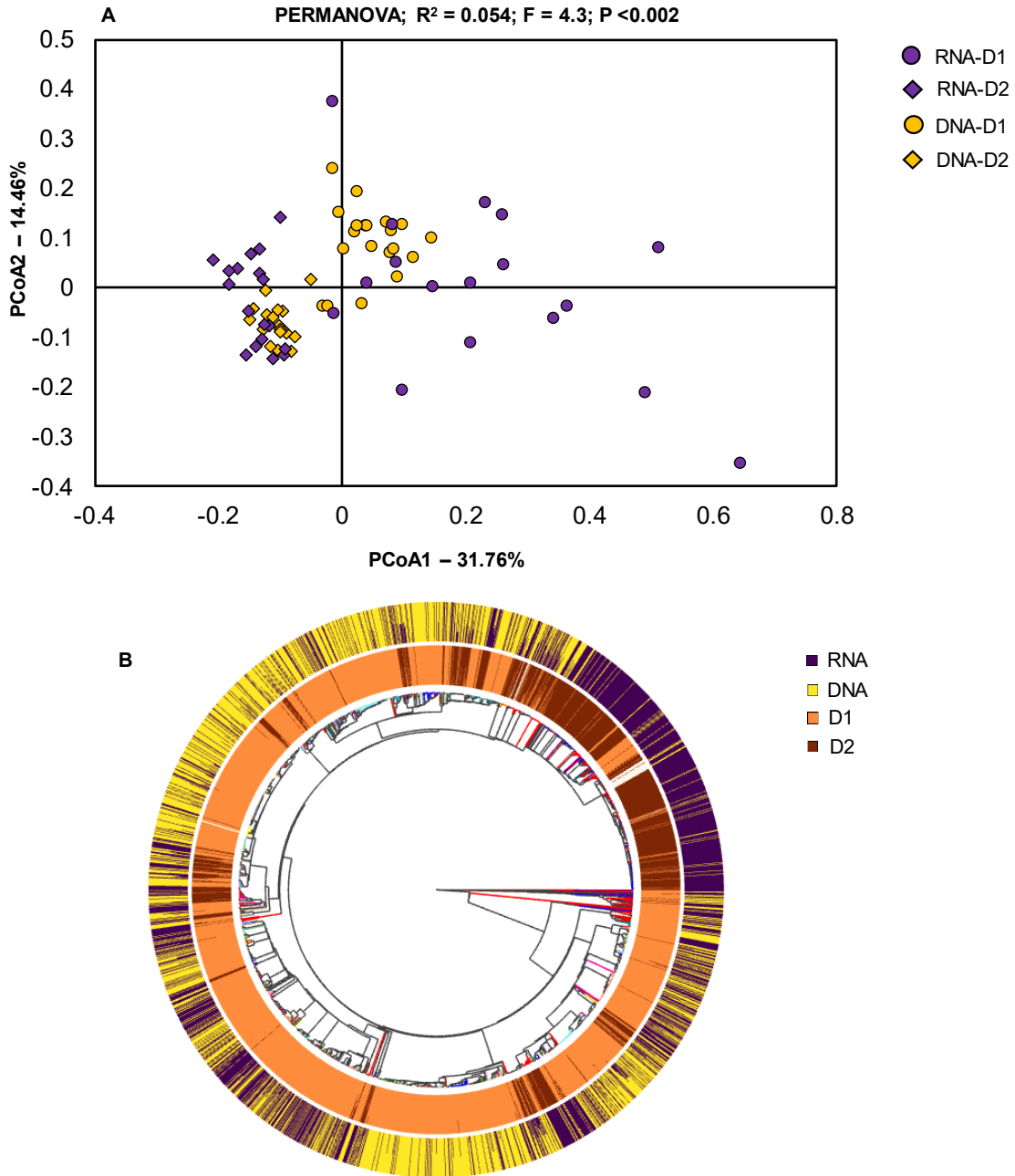
Figure 3.7 Comparison of fecal bacterial communities generated using DNA and RNA-based bacterial profiling (A) principal coordinate analysis (PCoA) using weighted UniFrac distance matrix in QIIME2 platform and PERMANOVA statistical test and (B) phylogenetic tree shows the phylogenetic similarity and distances of bacterial taxa identified from different genetic materials. The outer most bar plot represents different genetic materials and the inner bar plot represents sampling timepoints.
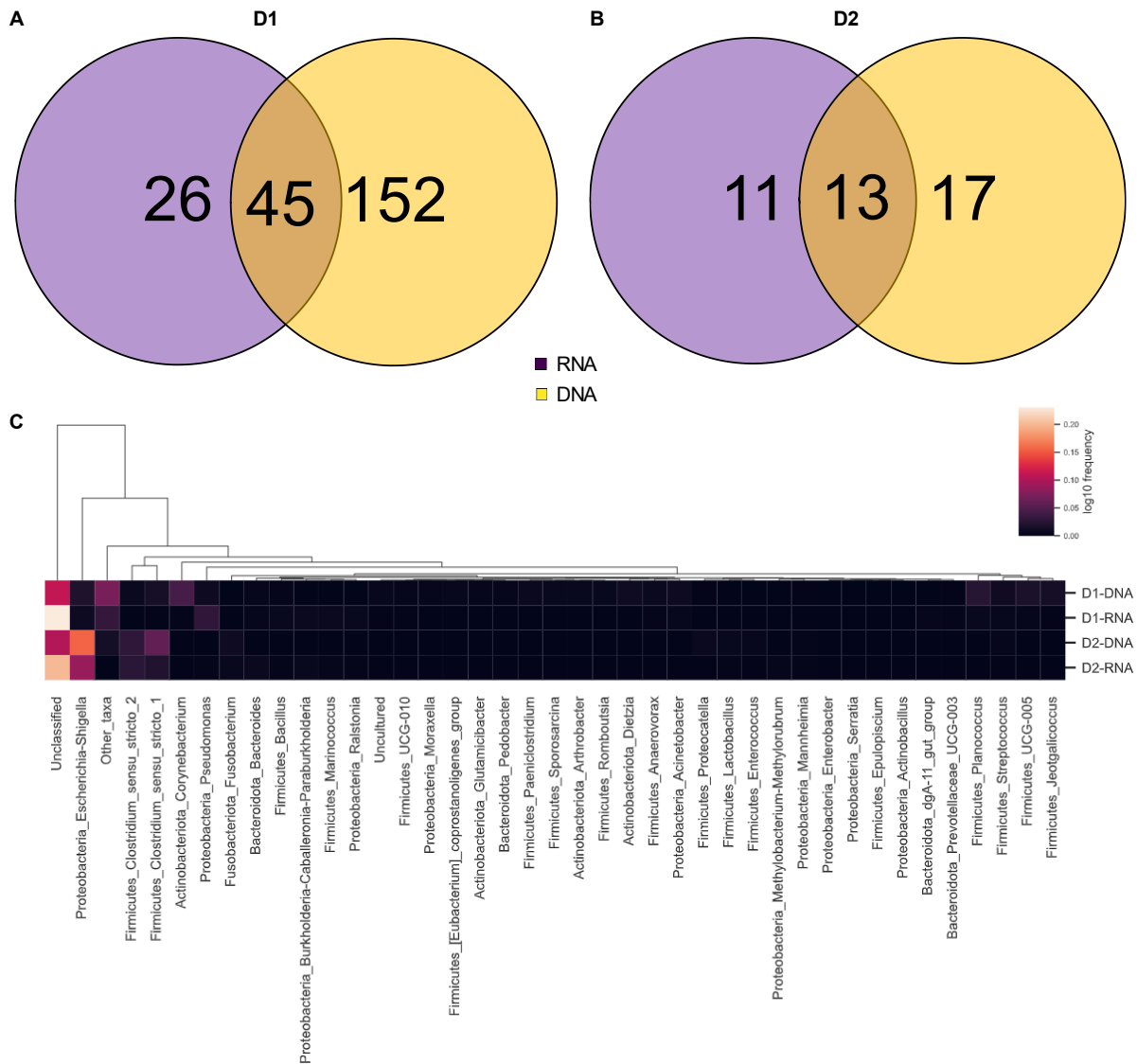
Figure 3.8 Taxonomic classification of fecal bacterial community between DNA and RNA-based amplicon sequencing approaches (A) Venn diagram represent the total number of unique and shared genera sequenced using different genetic materials in all calves (n=20) on day one (D1), (B) on day two (D2) and (C) heat map compared the relative abundance of forty (40) bacterial genera identified using two different genetic material used. Bacterial taxa are arranged from the highest relative abundance to lowest (phylum_genus).
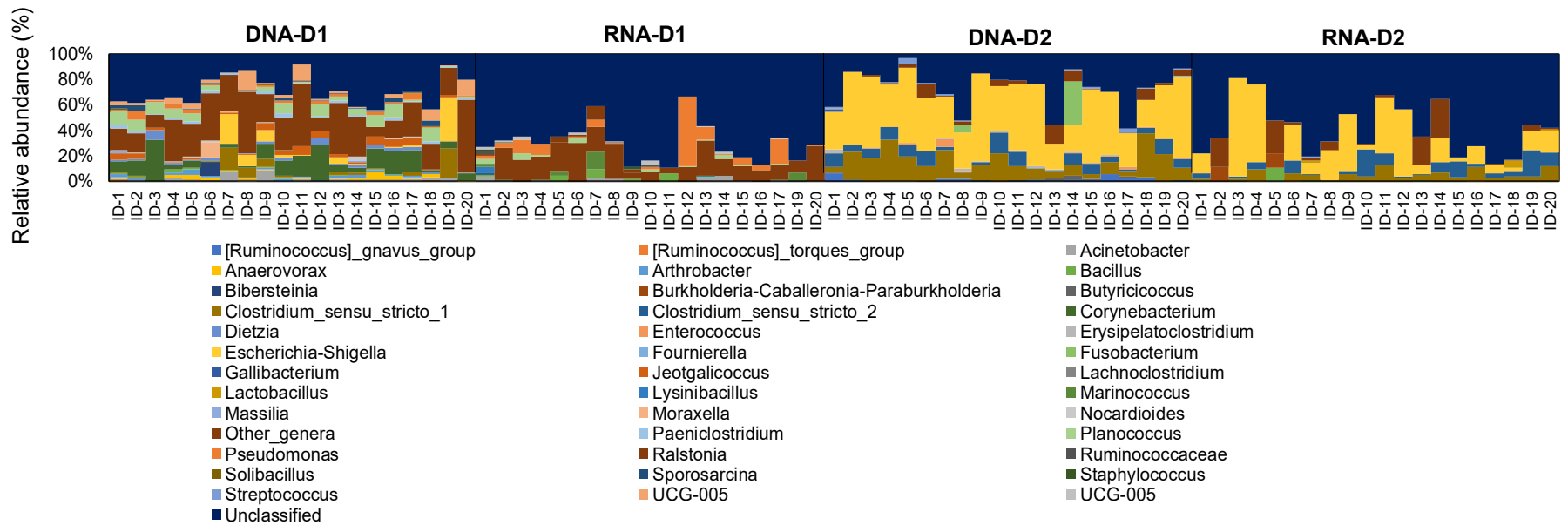
Figure 3.9 Relative abundance (for individual animals; n = 20) of the fecal bacterial communities between DNA and RNA-based amplicon sequencing approaches (genus level).
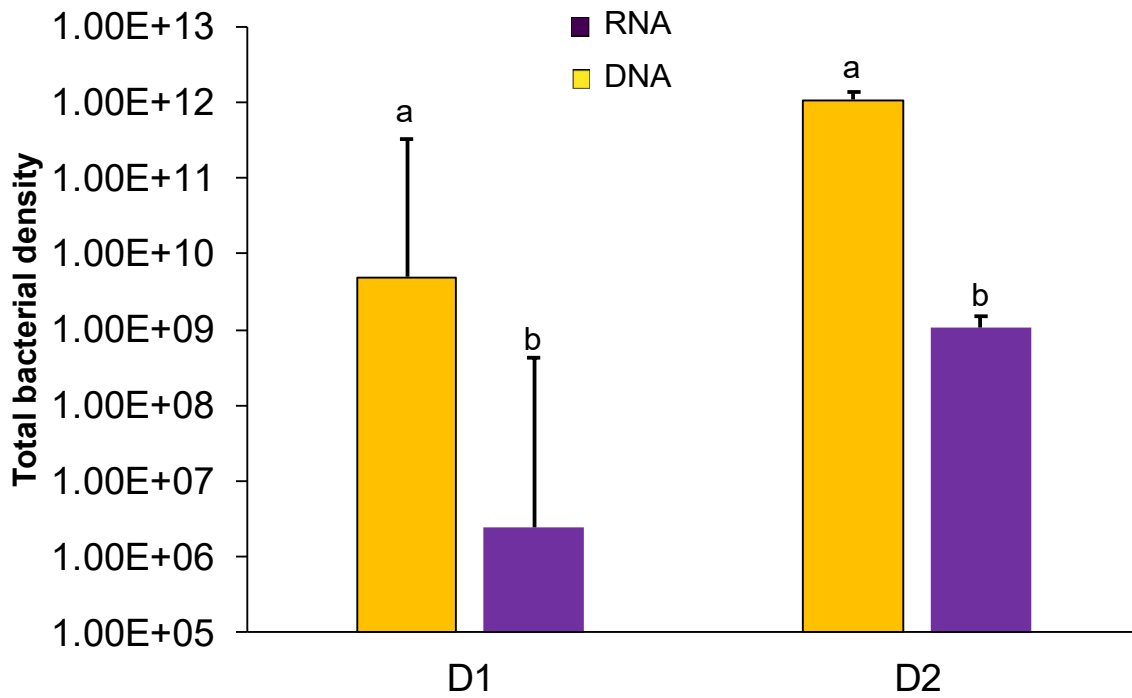
Figure 3.10 Total (DNA) and RNA-based bacterial densities (16S rRNA gene copy/swab) in rectal swab samples. Mean bacterial densities represent in bars and standard error mean (SEM) denoted by error bars. [a, b] means with different superscript are significantly different ($p<0.05$).
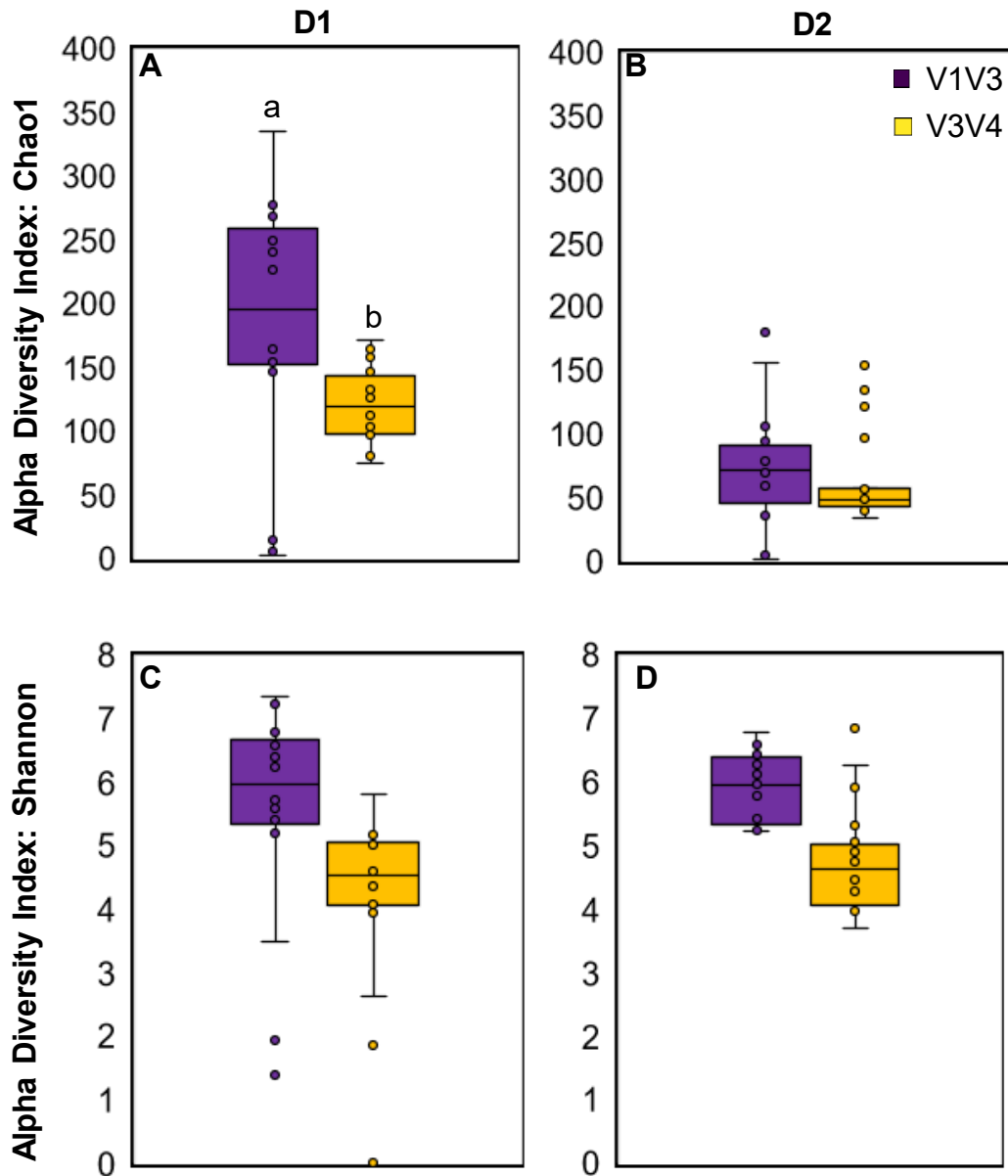
Figure 3.11 Oral bacterial diversity comparison (DNA-based amplicon sequencing) between V1V3 and V3V4 hypervariable regions using (A) alpha diversity index Chao 1 on day one (D1), (B) on day two (D2), (C) alpha diversity index Shannon on day one and (D) on day two; [a, b] means with different superscript are significantly different at $p<0.05$.
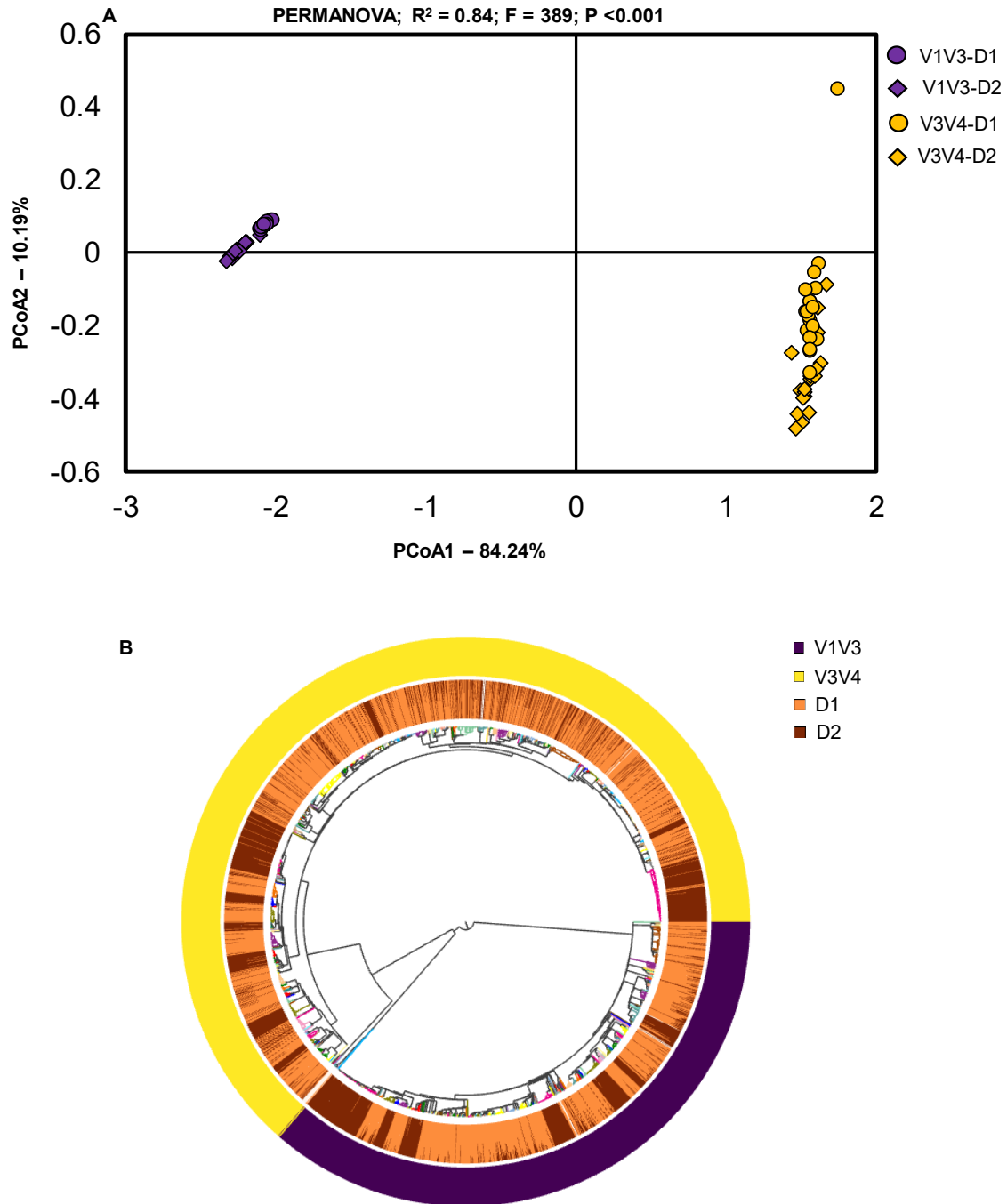
Figure 3.12 Comparison of oral bacterial communities generated by targeting V1V3 and V3V4 hypervariable regions (A) principle coordinate analysis (PCoA) using weighted UniFrac distance metrics in QIIME2 platform and PERMANOVA test and (B) phylogenetic tree shows the phylogenetic similarity and distances of bacterial taxa identified from different hypervariable regions. The outer most bar plot represents different hypervariable regions and the inner bar plot represents sampling timepoints.
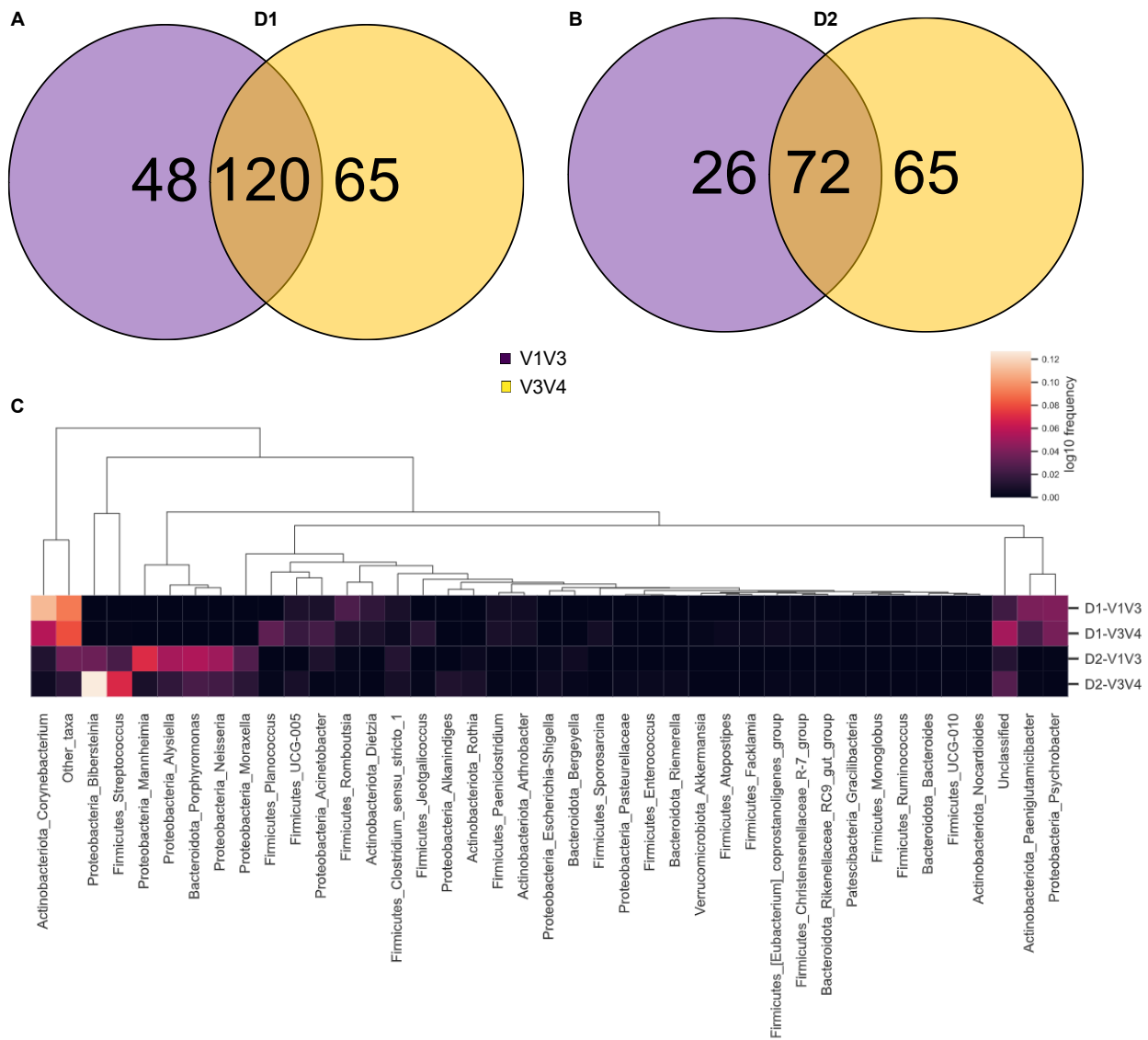
Figure 3.13 Taxonomic classification of oral bacterial community generated by targeting two hypervariable regions (V1V3 and V3V4) of the 16s rRNA marker gene (A) Venn diagrams represent the total number of unique and shared genera present in different regions in all calves (n=20) on day one, (B) on day two and (C) heat map compare the relative abundance of forty (40) bacterial genera identified using two primer pairs. Bacterial taxa are arranged from the highest relative abundance to lowest (phylum_genus).
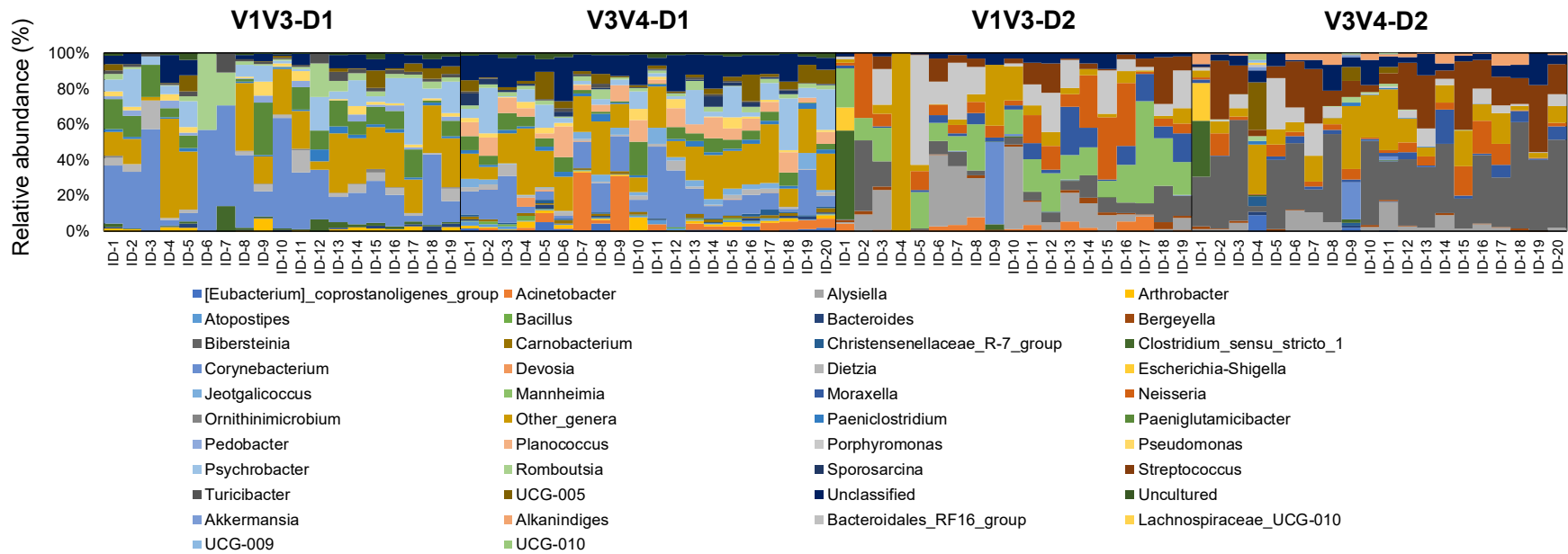
Figure 3.14 Relative abundance (for individual animals; n = 20) of the oral bacterial communities generated by targeting two hypervariable regions (V1V3 and V3V4) of the 16S rRNA marker gene (genus level).
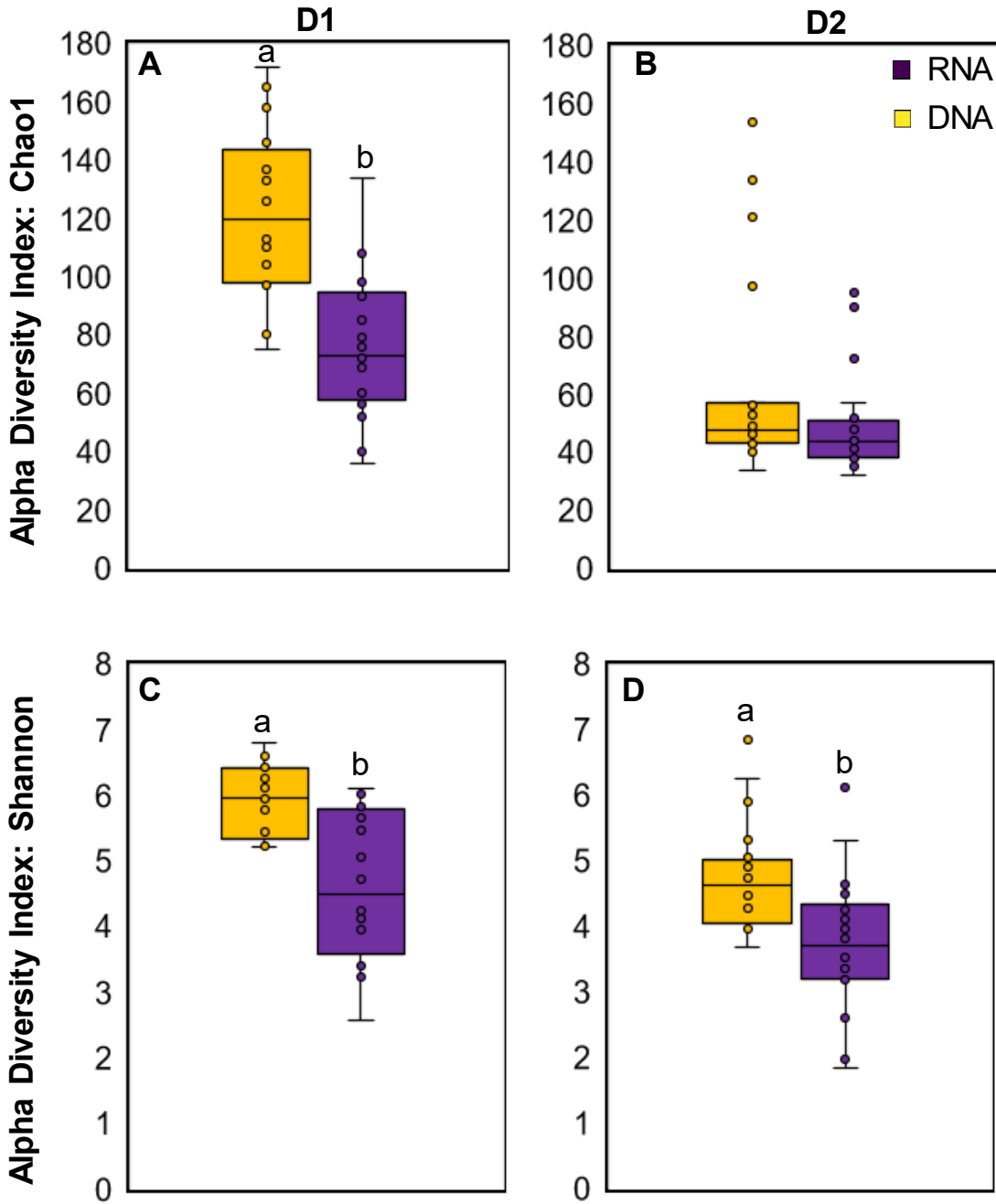
Figure 3.15 Oral bacterial diversity comparison between DNA and RNA-based amplicon sequencing of 16S rRNA marker gene targeting V3V4 region (A) alpha diversity index Chao 1 on day 1 (D1), (B) on day 2 (D2), (C) alpha diversity index Shannon on day one (D1) and (D) on day two (D2); [a, b] means with different superscript are significantly different at $p<0.05$.
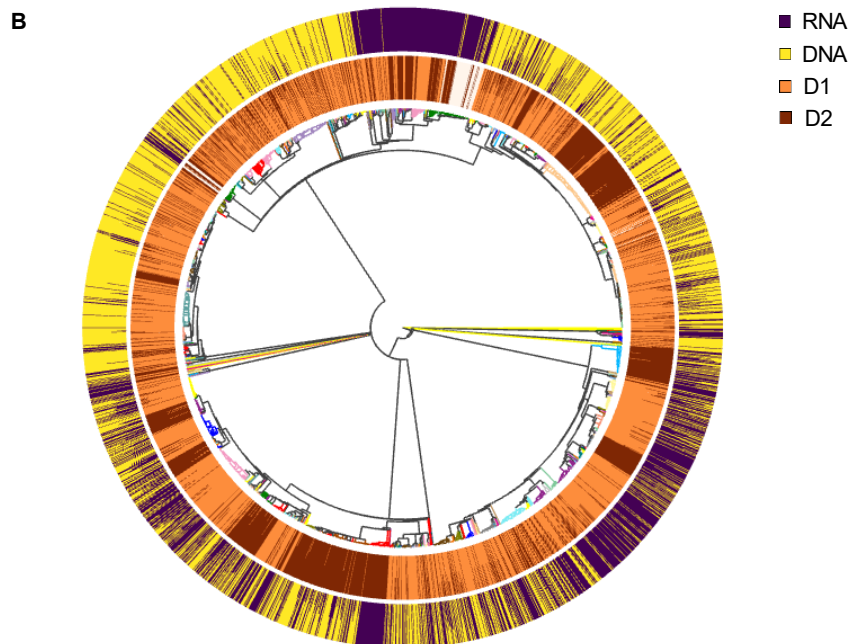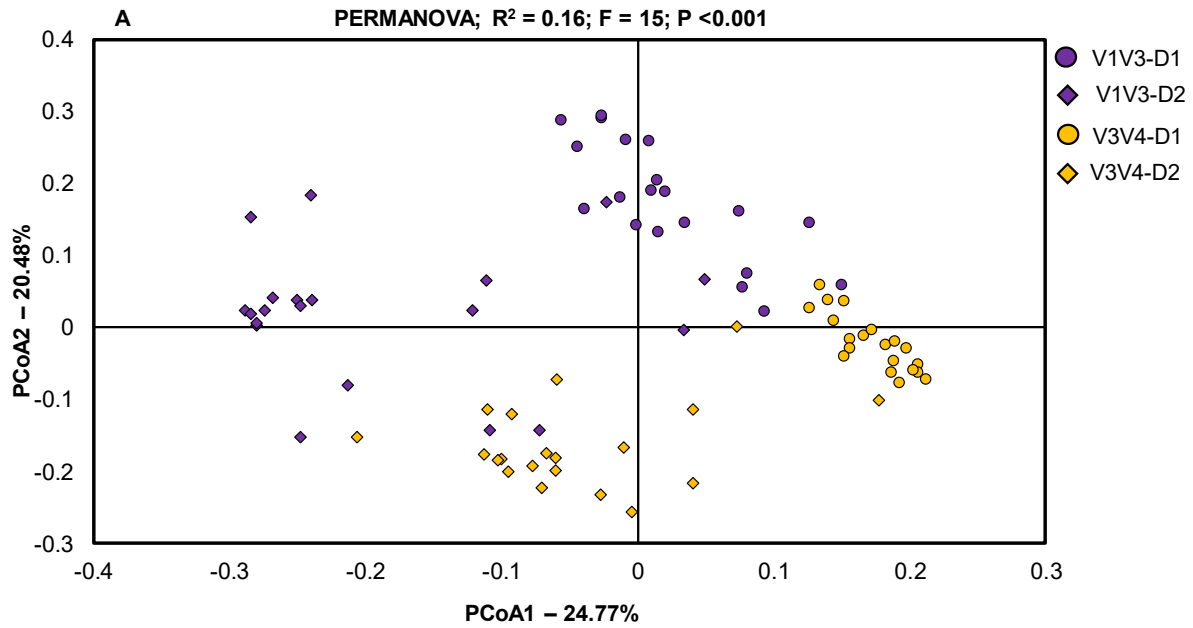
Figure 3.16 Comparison of oral bacterial communities generated using DNA and RNA-based bacterial profiling (A) principle coordinate analysis (PCoA) using weighted UniFrac distance matrix in QIIME2 platform and PERMANOVA statistical test and (B) phylogenetic tree shows the phylogenetic similarity and distances of bacterial taxa identified from different genetic materials. The outer most bar plot represents different genetic materials and the inner bar plot represents sampling timepoints.
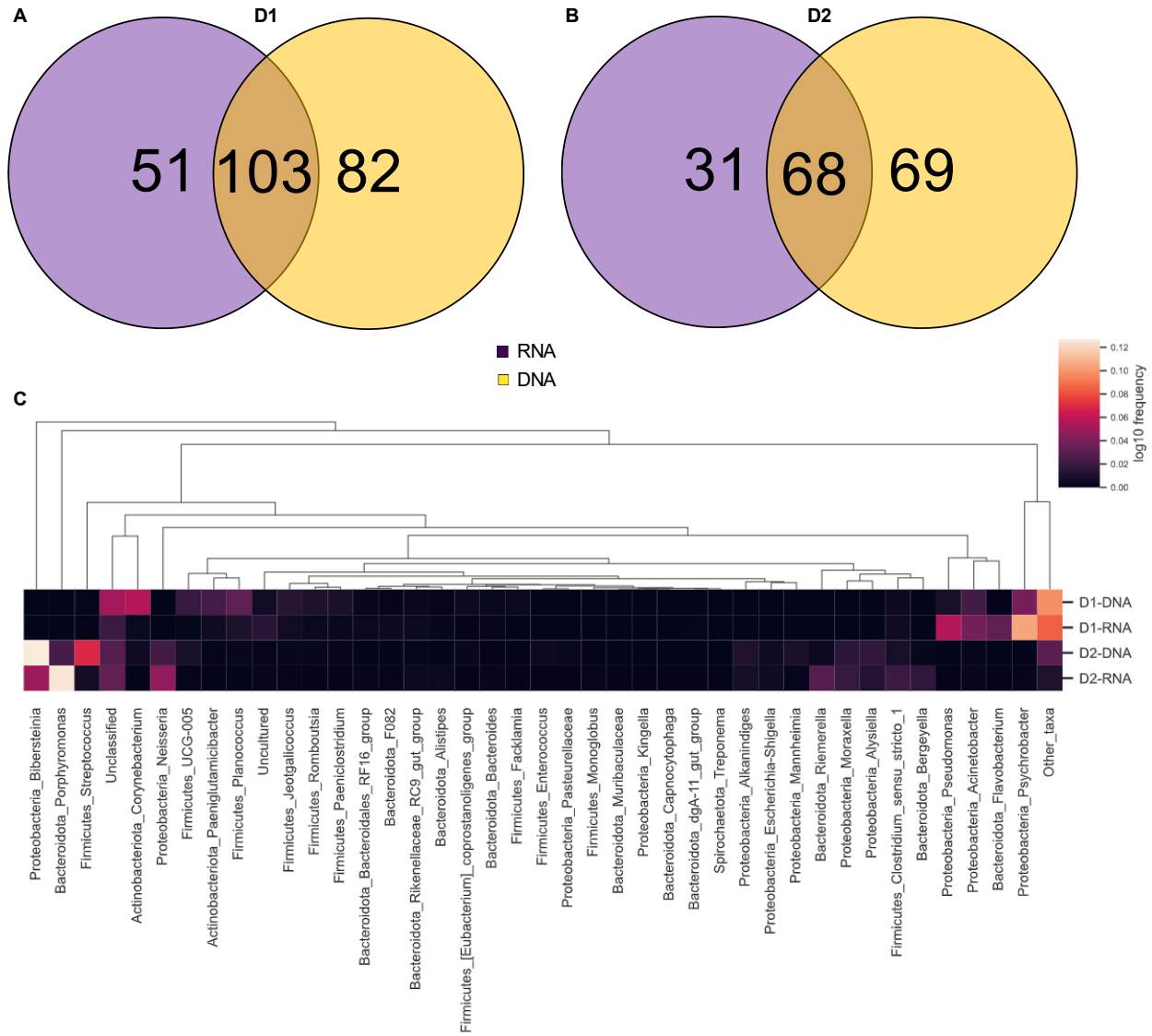
Figure 3.17 Taxonomic classification of oral bacterial community between DNA and RNA-based amplicon sequencing approaches (A) Venn diagram represent the total number of unique and shared genera sequenced using different genetic materials in all calves (n=20) on day one (D1), (B) on day two (D2) and (C) heat map compared the relative abundance of forty (40) bacterial genera identified using two different genetic material used. Bacterial taxa are arranged from the highest relative abundance to lowest (phylum_genus).
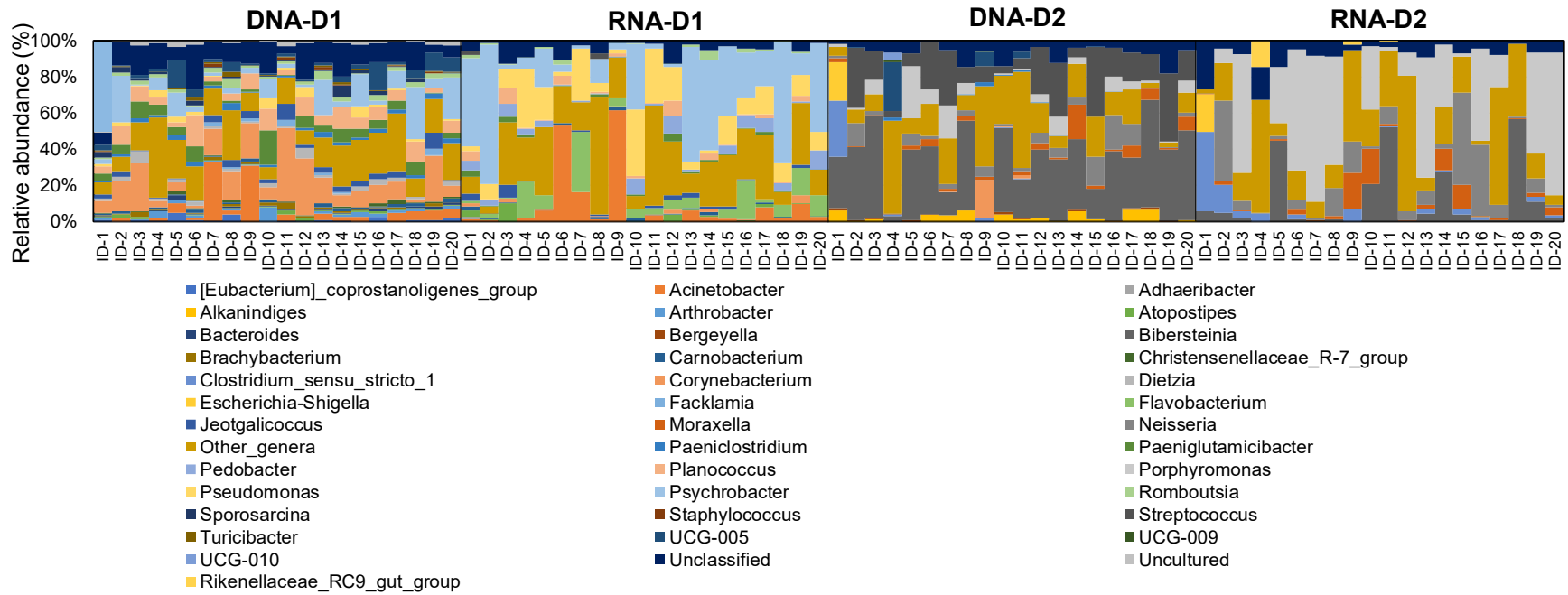
Figure 3.18 Relative abundance (for individual animals; n = 20) of the oral bacterial communities between DNA and RNA-based amplicon sequencing approaches (genus level).

Figure 3.19 Total bacterial density (DNA and RNA-based (16S rRNA gene copy/swab) in oral swab samples. Mean bacterial densities represent in bars and standard error mean (SEM) denoted by error bars. [a, b] means with different superscript are significantly different ($p<0.05$).
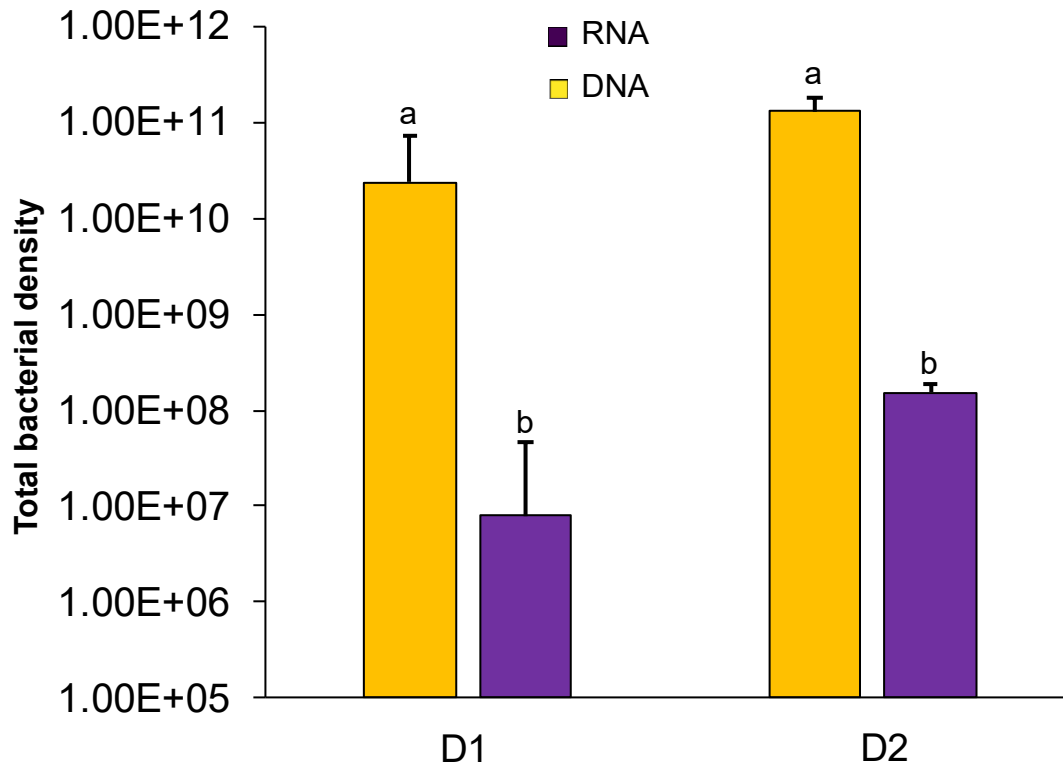
Chapter 4. General Discussion

4.1 Significance of the study

Host-microbial interactions play a major role in human and livestock species by affecting the growth and overall health of the host. The diverse microbial community colonizing the GIT during early life is critical for the development of the immune system and the maintenance of gut homeostasis and gut health (Nakandalage et al., 2023). However, early life microbial colonization is affected by various factors, leading to microbial dysbiosis and various negative health outcomes. Early life provides a window of opportunity to manipulate microbial communities (Robertson et al., 2019), suggesting that interventions during early life can be used to restore a perturbed microbial community and revert negative health outcomes (Mostafa et al., 2020; Barratt et al., 2022). However, the development of successful microbial interventions depends on the accurate identification of key beneficial microbiota during early life.

Due to the advanced next-generation sequencing and user-friendly bioinformatics pipelines, microbial community profiling has become popular among researchers (Callahan et al., 2017). However, microbial profiling using low microbial biomass samples has created debatable results (Rackaityte et al., 2020; Guzman et al., 2020; Stinson et al., 2020; Mishra et al., 2021). For example, previous studies reported that the identification of microbial community in the fetal environment was due to contaminations (Lauder et al., 2016; Lim et al., 2018; Malmuthuge and Griebel, 2018; De Goffau et al., 2019; Kennedy et al., 2021; Kennedy et al., 2023), suggesting profiling of low microbial community samples is challenging.

Appropriate molecular techniques, proper controls, and optimization of data analysis pipelines allow researchers to generate credible outcomes in microbial research, especially when using samples with low microbial biomass. Previous studies reported that the optimization of bioinformatics pipelines increased the credibility of microbial diversity and community profiling outcomes (Malmuthuge and Griebel, 2018; Rai et al., 2021). Our study revealed that optimization of denoising parameters increases the credibility of data generated using low microbial biomass samples. In the present study, we compared different hypervariable regions (V1V3 and V3V4) and genetic materials (DNA and RNA) to identify their impact on bacterial community composition. Based on the present findings, we can interpret that oral and fecal bacterial communities (low microbial biomass samples) can be profiled using the V1V3 region. In our study, we identified higher unclassified taxa when RNA-based sequencing was used for bacterial profiling. We

speculated that the bacteria community identified through RNA-based sequencing cannot be classified based on current databases, which suggests that the active bacterial community in newborn beef calves has not been well defined. Thus, future research should classify active bacterial community colonization in newborn beef calves. The knowledge generated through this study will provide clear guidelines to any researcher who wants to include a microbial analysis in their research.

4.2 Importance of optimizing denoising parameters

Open-source, user-friendly bioinformatics tools for microbial data analysis allow researchers to summarize and interpret high throughput data generated through NGS technologies. There are many bioinformatics tools that have been introduced to perform data analysis, including microbial diversity, composition, and phylogenetic analysis. Some bioinformatics tools not only perform data analysis but also conduct statistical analysis and data visualization. QIIME2 is one of the open-source, user-friendly bioinformatics tools for microbial data analysis, which can perform statistical analyses and data visualization (Mohsen et al., 2019). The QIIME2 bioinformatics pipeline contains numerous plugins with different algorithms to perform various microbial data analyses. Denoising is one of the most important and time-consuming plugins in microbial data analysis. During this step, researchers can perform quality filtering, remove chimeric sequences, and join paired-end reads (Callahan et al., 2016; Hall and Beiko, 2018; Estaki et al., 2020). However, the use of default settings in denoising is not suitable for all microbial analyses.

Generally, researchers use the mean quality scores to decide the denoising parameters. The use of the mean to decide the denoising parameters might include outliers in the data set. These extreme values (outliers) may misinterpret the data set (in this study the quality of the sequences to decide the truncation lengths). However, the use of the median quality score to decide denoising parameters provides an opportunity to avoid bias generated through outliers and represent the whole dataset to decide the denoising parameters. In our study, we used default denoising parameters in our analysis by deciding read truncation based on mean Phred score ≥25. Default denoising parameters, resulted in an extremely lower percentage of merged reads (~1%) after denoising. The overlapped base pair count between forward and reverse reads should be at least 20 bp (Callahan et al., 2016; Hall and Beiko, 2018). However, we could not achieve the recommended level of overlap when using default parameters to decide truncation length, suggesting the importance of re-evaluating sequence stats to identify approaches to improve data analysis.

When denoising parameters were optimized to maintain a minimum of 20 bp between forward and reverse reads, while maintaining a median Phred score ≥25 as the quality standard, it increased the number of merged reads significantly. As a result, we could see that there is an increase in the number of samples used for downstream analysis to assess bacterial diversity and taxonomic classification. During bacterial analysis, the inclusion of more samples in the downstream analysis increases our opportunities to assess the impact of treatments on microbial communities. It suggests that the sample size needs to be sufficient (statistical power) to get better interpretation during downstream analysis. Based on current findings we can interpret that optimized denoising parameters increase the accuracy of bacterial community profiling when using samples with low microbial biomass.

4.3 Use of appropriate molecular techniques to profile bacteria in low microbial biomass samples

Contaminations during microbial profiling lead to a misrepresentation of microbial diversity and composition. Contaminations can be introduced from sample collection to the end of data analysis (Eisenhofer et al., 2018; Kennedy et al., 2023). The use of appropriate controls allows the identification of contaminations and quality filtering steps can be used to remove identified contaminations. Contaminated taxa were identified irrespective of the hypervariable region and the genetic material used to profile the microbial community. However, a number of contaminated taxa observed in oral and fecal samples were lower in the microbial profiles generated by targeting the V1V3 region and using RNA.

Microbial community composition varied depending on the choice of the hypervariable region of the 16S rRNA gene and the choice of genetic materials to profile microbial communities. In the present study, we investigated the suitable region and genetic material to profile the oral and fecal bacterial community using low microbial biomass samples. Our study revealed a higher bacterial diversity and distinct bacterial community with phylogenetically different clusters in the V1V3 region compared to the V3V4 region while generating a lower level of contaminant ASVs. Therefore, we conclude that the use of the V1V3 region increases the accuracy of bacterial analysis when using low microbial biomass samples. When looking at the suitable genetic material to profile low microbial biomass samples, RNA-based profiling provides higher diversity and lower unclassified genera in oral samples of newborn beef calves, suggesting that it is suitable for studying oral bacterial diversity and community composition. The fecal bacterial community of newborn beef calves can be studied using DNA-based microbial profiling due to the presence of higher unclassified

genera in RNA-based bacterial profiling. However, RNA-based bacterial profiling is suitable to study fecal bacterial diversity in newborn beef calves.

4.4 Future directions

Our study provided clear and specific knowledge/guidelines on profiling bacterial using low microbial biomass samples. This study can be used as a basis to optimize denoising parameters during data analysis in low microbial biomass samples and to identify true bacterial communities with appropriate controls. Thus, our findings will help researchers to reduce time on bacterial data analysis during denoising, which is one of the important and time-consuming processes. Moreover, this study provides fundamental knowledge on the choice of the hypervariable region of the 16S rRNA gene and genetic materials to profile bacteria. Limiting our RNA-based analysis only to the V3V4 region is one of the limitations of our study. Therefore, we will not be able to accurately evaluate the impact of using RNA on bacterial profiling. It is important to compare RNA-based bacterial community profiles using different variable regions to identify variations in bacterial diversity, composition, and taxonomic classification. Moreover, this study will provide more opportunities for calf researchers to collect samples and conduct bacterial community profiling of newborn calves using swab samples to profile early-life bacteria over a period of time to assess bacterial dynamics during early life, disease progression, and treatments.

4.5 References

Barratt MJ, Ahmed T, Gordon JI. Gut microbiome development and childhood undernutrition. Cell Host Microbe. 2022;30(5):617-26.

Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11(12):2639-43.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.

De Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, et al. Human placenta has no microbiome but can contain potential pathogens. Nature. 2019;572(7769):329-34.

Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol. 2019;27(2):105-17.

Estaki M, Jiang L, Bokulich NA, McDonald D, Gonzalez A, Kosciolek T, et al. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. Curr Protoc Bioinformatics. 2020;70(1):e100.

Guzman CE, Wood JL, Egidi E, White-Monsant AC, Semenec L, Grommen SVH, et al. A pioneer calf foetus microbiome. Sci Rep. 2020;10(1):17712.

Hall M, Beiko RG. 16S rRNA gene analysis with QIIME2. Microbiome analysis: methods and protocols. 2018:113-29.

Kennedy KM, de Goffau MC, Perez-Munoz ME, Arrieta MC, Backhed F, Bork P, et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. Nature. 2023;613(7945):639-49.

Kennedy KM, Gerlach MJ, Adam T, Heimesaat MM, Rossi L, Surette MG, et al. Fetal meconium does not have a detectable microbiota before birth. Nat Microbiol. 2021;6(7):865-73.

Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, et al. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. Microbiome. 2016;4(1):29.

Lim ES, Rodriguez C, Holtz LR. Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. Microbiome. 2018;6(1):87.

Malmuthuge N, Griebel PJ. Fetal environment and fetal intestine are sterile during the third trimester of pregnancy. Vet Immunol Immunopathol. 2018;204:59-64.

Mishra A, Lai GC, Yao LJ, Aung TT, Shental N, Rotter-Maskowitz A, et al. Microbial exposure during early human development primes fetal immune cells. Cell. 2021;184(13):3394-409 e20.

Mohsen A, Park J, Chen YA, Kawashima H, Mizuguchi K. Impact of quality trimming on the efficiency of reads joining and diversity analysis of Illumina paired-end reads in the context of QIIME1 and QIIME2 microbiome analysis frameworks. BMC Bioinformatics. 2019;20(1):581.

Mostafa I, Nahar NN, Islam MM, Huq S, Mustafa M, Barratt M, et al. Proof-of-concept study of the efficacy of a microbiota-directed complementary food formulation (MDCF) for treating moderate acute malnutrition. BMC Public Health. 2020;20(1):242.

Nakandalage R, Guan L L, Malmuthuge N, et al. Microbial Interventions to Improve Neonatal Gut Health Microorganisms. 2023; 11, no. 5: 1328.

Rackaityte E, Halkias J, Fukui EM, Mendoza VF, Hayzelden C, Crawford ED, et al. Viable bacterial colonization is highly limited in the human intestine in utero. Nat Med. 2020;26(4):599-607.

Rai SN, Qian C, Pan J, Rai JP, Song M, Bagaitkar J, et al. Microbiome data analysis with applications to pre-clinical studies using QIIME2: Statistical considerations. Genes Dis. 2021;8(2):215-23.

Robertson RC, Manges AR, Finlay BB, Prendergast AJJTim. The human microbiome and child growth– first 1000 days and beyond. 2019;27(2):131-47.

Stinson L, Hallingstrom M, Barman M, Viklund F, Keelan J, Kacerovsky M, et al. Comparison of Bacterial DNA Profiles in Mid-Trimester Amniotic Fluid Samples From Preterm and Term Deliveries. Front Microbiol. 2020;11:415.

Bibliography (All sources used)

Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. Sci. transl. med. 2014 May 21;6(237):237ra65-.

Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, et al. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. Msphere. 2021;6(1):e01202-20.

Al Nabhani Z, Eberl G. Imprinting of the immune system by the microbiota early in life. Mucosal Immunol. 2020;13(2):183-9.

Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2017;2(2).

Angly FE, Dennis PG, Skarshewski A, Vanwonterghem I, Hugenholtz P, Tyson GW. CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. Microbiome. 2014 ;2(1):1-3.

Araujo G, Yunta C, Terré M, Mereu A, Ipharraguerre I, Bach A. Intestinal permeability and incidence of diarrhea in newborn calves. J. dairy sci. 2015 ;98(10):7309-17.

Ballenghien M, Faivre N, Galtier N. Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. BMC Biol. 2017;15(1):25.

Ballou AL, Ali RA, Mendoza MA, Ellis JC, Hassan HM, Croom WJ, et al. Development of the Chick Microbiome: How Early Exposure Influences Future Microbial Diversity. Front Vet Sci. 2016;3:2.

Barden M, Richards-Rios P, Ganda E, Lenzi L, Eccles R, Neary J, et al. Maternal influences on oral and faecal microbiota maturation in neonatal calves in beef and dairy production systems. Anim Microbiome. 2020;2(1):31.

Barratt MJ, Ahmed T, Gordon JI. Gut microbiome development and childhood undernutrition. Cell Host Microbe. 2022;30(5):617-26.

Bäumler AJ, Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. Nature. 2016 ;535(7610):85-93.

Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform. 2021;22(1):178-93.

Bischoff SC. 'Gut health': a new objective in medicine?. BMC medicine. 2011 ;9:1-4.

Bleidorn C, Bleidorn C. Sequencing techniques. Phylogenomics: An Introduction. 2017:43-60.

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018 ;6(1):1-7.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852-7.

Caballero-Flores G, Pickard JM, Nunez G. Microbiota-mediated colonization resistance: mechanisms and regulation. Nat Rev Microbiol. 2022.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA. QIIME allows analysis of high-throughput community sequencing data. Nat methods. 2010 ;7(5):335-6.

Carlsen T, Aas AB, Lindner D, Vrålstad T, Schumacher T, Kauserud H. Don't make a mista (g) ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies?. Fungal Ecol. 2012;5(6):747-9.

Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. microbiol methods. 2007;69(2):330-9.

Chen C, Khaleel SS, Huang H, Wu CH. Software for pre-processing Illumina next-generation sequencing short read sequences. Source code biol. med. 2014;9(1):1-1.

Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MC, Wong MC, Ng SC, Chan FK, Chan PK. Impact of preservation method and 16S rRNA hypervariable region on gut microbiota profiling. Msystems. 2019 Feb 26;4(1):e00271-18.

Collado MC, Cernada M, Neu J, Perez-Martinez G, Gormaz M, Vento M. Factors influencing gastrointestinal tract and microbiota immune interaction in preterm infants. Pediatr Res. 2015;77(6):726-31.

Constantinides MG, Link VM, Tamoutounour S, Wong AC, Perez-Chaparro PJ, Han SJ, et al. MAIT cells are imprinted by the microbiota in early life and promote tissue repair. Science. 2019;366(6464).

Correa-Fiz F, Goncalves Dos Santos JM, Illas F, Aragon V. Antimicrobial removal on piglets promotes health and higher bacterial diversity in the nasal microbiota. Sci Rep. 2019;9(1):6545.

Crosby CM, Kronenberg M. Tissue-specific functions of invariant natural killer T cells. Nat. Rev. Immunol. 2018 ;18(9):559-74.

Danzeisen JL, Clayton JB, Huang H, Knights D, McComb B, Hayer SS, et al. Temporal Relationships Exist Between Cecum, Ileum, and Litter Bacterial Microbiomes in a Commercial Turkey Flock, and Subtherapeutic Penicillin Treatment Impacts Ileum Bacterial Community Establishment. Front Vet Sci. 2015;2:56.

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(1):226.

De Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, et al. Human placenta has no microbiome but can contain potential pathogens. Nature. 2019;572(7769):329-34.

De Vrieze J, Pinto AJ, Sloan WT, Ijaz UZ. The active microbial community more accurately reflects the anaerobic digestion process: 16S rRNA (gene) sequencing as a predictive tool. Microbiome. 2018;6(1):63.

Doyle RM, Alber DG, Jones HE, Harris K, Fitzgerald F, Peebles D, et al. Term and preterm labour are associated with distinct microbial community structures in placental membranes which are independent of mode of delivery. Placenta. 2014;35(12):1099-101.

Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol. 2019;27(2):105-17.

Ellermann M, Carr JS, Fodor AA, Arthur JC, Carroll IM. Characterizing and functionally defining the gut microbiota: methodology and implications. InThe microbiota in gastrointestinal pathophysiology 2017 (pp. 15-25). Academic Press.

Estaki M, Jiang L, Bokulich NA, McDonald D, Gonzalez A, Kosciolek T, et al. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. Curr Protoc Bioinformatics. 2020;70(1):e100.

Fischer AJ, Song Y, He Z, Haines DM, Guan LL, Steele MA. Effect of delaying colostrum feeding on passive transfer and intestinal bacterial colonization in neonatal male Holstein calves. J Dairy Sci. 2018;101(4):3099-109.

Fouhse JM, Yang K, More-Bayona J, Gao Y, Goruk S, Plastow G, et al. Neonatal Exposure to Amoxicillin Alters Long-Term Immune Response Despite Transient Effects on Gut-Microbiota in Piglets. Front Immunol. 2019;10:2059.

Fraher MH, O'Toole PW, Quigley EM. Techniques used to characterize the gut microbiota: a guide for the clinician. Nat Rev Gastroenterol Hepatol. 2012;9(6):312-22.

Furman O, Shenhav L, Sasson G, Kokou F, Honig H, Jacoby S, Hertz T, Cordero OX, Halperin E, Mizrahi I. Stochasticity constrained by deterministic effects of diet and age drive rumen microbiome assembly dynamics. Nat. commun. 2020 ;11(1):1904.

Geleto A, Chojenta C, Taddele T, Loxton D. Association between maternal mortality and caesarean section in Ethiopia: a national cross-sectional study. BMC Pregnancy and childbirth. 2020 ;20:1-1.

Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. Science. 2016 ;352(6285):539-44.

Ghaffari MH, Sadri H, Steinhoff-Wagner J, Hammon HM, Sauerwein H. Effects of colostrum feeding on the mRNA abundance of genes related to toll-like receptors, key antimicrobial defense molecules, and tight junctions in the small intestine of neonatal dairy calves. J Dairy Sci. 2021;104(9):10363-73.

Ghosh S, Whitley CS, Haribabu B, Jala VR. Regulation of Intestinal Barrier Function by Microbial Metabolites. Cell Mol Gastroenterol Hepatol. 2021;11(5):1463-82.

Gillilland GM, Young VB, Huffnagle GB. Gastrointestinal Microbial Ecology with Perspectives on Health and Disease. Editor(s): Leonard R. Johnson, Fayez K. Ghishan, Jonathan D. Kaunitz, Juanita L. Merchant, Hamid M. Said, Jackie D. Wood, Physiology of the Gastrointestinal Tract (Fifth Edition) 2012 (pp 1119-1134) Academic Press.

Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog. 2016;8:24.

Golob JL, Margolis E, Hoffman NG, Fredricks DN. Evaluating the accuracy of amplicon-based microbiome computational pipelines on simulated human gut microbial communities. BMC Bioinformatics. 2017;18(1):283.

Gonzalez-Gonzalez M, Diaz-Zepeda C, Eyzaguirre-Velasquez J, Gonzalez-Arancibia C, Bravo JA, Julio-Pieper M. Investigating Gut Permeability in Animal Models of Disease. Front Physiol. 2018;9:1962.

Guevarra RB, Lee JH, Lee SH, Seok MJ, Kim DW, Kang BN, et al. Piglet gut microbial shifts early in life: causes and effects. J Anim Sci Biotechnol. 2019;10:1.

Guzman CE, Wood JL, Egidi E, White-Monsant AC, Semenec L, Grommen SVH, et al. A pioneer calf foetus microbiome. Sci Rep. 2020;10(1):17712.

Hall M, Beiko RG. 16S rRNA gene analysis with QIIME2. Microbiome analysis: methods and protocols. 2018:113-29.

Hansson J, Bosco N, Favre L, Raymond F, Oliveira M, Metairon S, Mansourian R, Blum S, Kussmann M, Benyacoub J. Influence of gut microbiota on mouse B2 B cell ontogeny and function. Mol. immunol. 2011;48(9-10):1091-101.

Heida FH, Kooi EMW, Wagner J, Nguyen TY, Hulscher JBF, van Zoonen A, et al. Weight shapes the intestinal microbiome in preterm infants: results of a prospective observational study. BMC Microbiol. 2021;21(1):219.

Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME j. 2011 ;5(10):1571-9.

Ho NT, Li F, Lee-Sarwar KA, Tun HM, Brown BP, Pannaraj PS, Bender JM, Azad MB, Thompson AL, Weiss ST, Azcarate-Peril MA. Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. Nat. commun. 2018;9(1):4169.

Hornef MW, Torow N. 'Layered immunity'and the 'neonatal window of opportunity'–timed succession of non-redundant phases to establish mucosal host–microbial homeostasis after birth. Immunology. 2020;159(1):15-25.

Jamuar SS, D'Gama AM, Walsh CA. Somatic mosaicism and neurological diseases. InGenomics, circuits, and pathways in clinical neuropsychiatry 2016 (pp. 179-199). Academic Press.

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):5029.

Jovel J, Nimaga A, Jordan T, O'Keefe S, Patterson J, Thiesen A, et al. Metagenomics Versus Metatranscriptomics of the Murine Gut Microbiome for Assessing Microbial Metabolism During Inflammation. Front Microbiol. 2022;13:829378.

Kameoka S, Motooka D, Watanabe S, Kubo R, Jung N, Midorikawa Y, et al. Benchmark of 16S rRNA gene amplicon sequencing using Japanese gut microbiome data from the V1-V2 and V3-V4 primer sets. BMC Genomics. 2021;22(1):527.

Karstens L, Asquith M, Caruso V, Rosenbaum JT, Fair DA, Braun J, et al. Community profiling of the urinary microbiota: considerations for low-biomass samples. Nat Rev Urol. 2018;15(12):735-49.

Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput Biol. 2012;8(10):e1002743.

Kennedy KM, de Goffau MC, Perez-Munoz ME, Arrieta MC, Backhed F, Bork P, et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. Nature. 2023;613(7945):639-49.

Kennedy KM, Gerlach MJ, Adam T, Heimesaat MM, Rossi L, Surette MG, et al. Fetal meconium does not have a detectable microbiota before birth. Nat Microbiol. 2021;6(7):865-73.

Kim ET, Lee SJ, Kim TY, Lee HG, Atikur RM, Gu BH, et al. Dynamic Changes in Fecal Microbial Communities of Neonatal Dairy Calves by Aging and Diarrhea. Animals (Basel). 2021;11(4).

Kim M, Morrison M, Yu Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. J Microbiol Methods. 2011;84(1):81-7.

Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nature Reviews Microbiology. 2018;16(7):410-22.

Kogut MH, Arsenault RJ. Editorial: Gut Health: The New Paradigm in Food Animal Production. Front Vet Sci. 2016;3:71.

Lagier JC, Hugon P, Khelaifia S, Fournier PE, La Scola B, Raoult D. The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. Clin Microbiol Rev. 2015;28(1):237-64.

Lamouse-Smith ES, Tzeng A, Starnbach MN. The intestinal flora is required to support antibody responses to systemic immunization in infant and germ free mice. PLoS One. 2011;6(11):e27662.

Larsson AJM, Stanley G, Sinha R, Weissman IL, Sandberg R. Computational correction of index switching in multiplexed sequencing libraries. Nat Methods. 2018;15(5):305-7.

Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, et al. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. Microbiome. 2016;4(1):29.

Li F, Henderson G, Sun X, Cox F, Janssen PH, Guan le L. Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches. Front Microbiol. 2016;7:987.

Li F, Li C, Chen Y, Liu J, Zhang C, Irving B, Fitzsimmons C, Plastow G, Guan LL. Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. Microbiome. 2019 ;7:1-7.

Li N, Huang S, Jiang L, Wang W, Li T, Zuo B, et al. Differences in the Gut Microbiota Establishment and Metabolome Characteristics Between Low- and Normal-Birth-Weight Piglets During Early-Life. Front Microbiol. 2018;9:1798.

Li Y, Xia S, Jiang X, Feng C, Gong S, Ma J, et al. Gut Microbiota and Diarrhea: An Updated Review. Front Cell Infect Microbiol. 2021;11:625210.

Lim ES, Rodriguez C, Holtz LR. Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. Microbiome. 2018;6(1):87.

Litvak Y, Byndloss MX, Bäumler AJ. Colonocyte metabolism shapes the gut microbiota. Science. 2018;362(6418):eaat9076.

Liu Y, Li X, Zhu B, Zhao H, Ai Q, Tong Y, et al. Midtrimester amniotic fluid from healthy pregnancies has no microorganisms using multiple methods of microbiologic inquiry. Am J Obstet Gynecol. 2020;223(2):248 e1- e21.

Loh L, Ivarsson MA, Michaëlsson J, Sandberg JK, Nixon DF. Invariant natural killer T cells developing in the human fetus accumulate and mature in the small intestine. Mucosal immunol. 2014;7(5):1233-43.

Louca S, Doebeli M, Parfrey LW. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome. 2018;6(1):41.

Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One. 2012;7(2):e30087.2018;9:2755.

Ma J, Li Z, Zhang W, Zhang C, Zhang Y, Mei H, et al. Comparison of gut microbiota in exclusively breast-fed and formula-fed babies: a study of 91 term infants. Sci Rep. 2020;10(1):15792.

Ma T, Villot C, Renaud D, Skidmore A, Chevaux E, Steele M, et al. Linking perturbations to temporal changes in diversity, stability, and compositions of neonatal calf gut microbiota: prediction of diarrhea. ISME J. 2020;14(9):2223-35.

Malmuthuge N, Griebel PJ, Guan le L. The Gut Microbiome and Its Potential Role in the Development and Function of Newborn Calf Gastrointestinal Tract. Front Vet Sci. 2015;2:36.

Malmuthuge N, Griebel PJ. Fetal environment and fetal intestine are sterile during the third trimester of pregnancy. Vet Immunol Immunopathol. 2018;204:59-64.

Malmuthuge N, Guan LL. Understanding the gut microbiome of dairy calves: Opportunities to improve early-life gut health. J Dairy Sci. 2017;100(7):5996-6005.

Mardis ER. DNA sequencing technologies: 2006–2016. Nat. protoc. 2017 ;12(2):213-8.

Marizzoni M, Gurry T, Provasi S, Greub G, Lopizzo N, Ribaldi F, et al. Comparison of Bioinformatics Pipelines and Operating Systems for the Analyses of 16S rRNA Gene Amplicon Sequences in Human Fecal Samples. Front Microbiol. 2020;11:1262.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal. 2011 ;17(1):10-2.

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(1):35.

Maynard CL, Elson CO, Hatton RD, Weaver CT. Reciprocal interactions of the intestinal microbiota and immune system. Nature. 2012 ;489(7415):231-41.

McElroy SJ, Prince LS, Weitkamp JH, Reese J, Slaughter JC, Polk DB. Tumor necrosis factor receptor 1-dependent depletion of mucus in immature small intestine: a potential role in neonatal necrotizing enterocolitis. Am J Physiol Gastrointest Liver Physiol. 2011;301(4):G656-66.

McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217.

Minich JJ, Zhu Q, Janssen S, Hendrickson R, Amir A, Vetter R, Hyde J, Doty MM, Stillwell K, Benardini J, Kim JH. KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. MSystems. 2018 ;3(3):e00218-17.

Mishra A, Lai GC, Yao LJ, Aung TT, Shental N, Rotter-Maskowitz A, et al. Microbial exposure during early human development primes fetal immune cells. Cell. 2021;184(13):3394-409 e20.

Mohsen A, Park J, Chen YA, Kawashima H, Mizuguchi K. Impact of quality trimming on the efficiency of reads joining and diversity analysis of Illumina paired-end reads in the context of QIIME1 and QIIME2 microbiome analysis frameworks. BMC Bioinformatics. 2019;20(1):581.

Mostafa I, Nahar NN, Islam MM, Huq S, Mustafa M, Barratt M, et al. Proof-of-concept study of the efficacy of a microbiota-directed complementary food formulation (MDCF) for treating moderate acute malnutrition. BMC Public Health. 2020;20(1):242.

Motley ST, Picuri JM, Crowder CD, Minich JJ, Hofstadler SA, Eshoo MW. Improved multiple displacement amplification (iMDA) and ultraclean reagents. BMC genomics. 2014 ;15:1-0.

Mulder IE, Schmidt B, Stokes CR, Lewis M, Bailey M, Aminov RI, et al. Environmentally-acquired bacteria influence microbial diversity and natural innate immune responses at gut surfaces. BMC Biol. 2009;7:79.

Munshi A, Sharma V. Omics and edible vaccines. InOmics Technologies and Bio-Engineering 2018 (pp. 129-141). Academic Press.

Nagpal R, Tsuji H, Takahashi T, Kawashima K, Nagata S, Nomoto K, et al. Sensitive Quantitative Analysis of the Meconium Bacterial Microbiota in Healthy Term Infants Born Vaginally or by Cesarean Section. Front Microbiol. 2016;7:1997.

Nakandalage R, Guan L L, Malmuthuge N, et al. Microbial Interventions to Improve Neonatal Gut Health. Microorganisms. 2023; 11, no. 5: 1328.

Nguyen NH, Smith D, Peay K, Kennedy P. Parsing ecological signal from noise in next generation amplicon sequencing. New Phytologist. 2015 ;205(4):1389-93.

Nilsson RH, Ryberg M, Abarenkov K, Sjokvist E, Kristiansson E. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. FEMS Microbiol Lett. 2009;296(1):97-101.

Nuzhat S, Hasan SMT, Palit P, Islam MR, Mahfuz M, Islam MM, et al. Effects of probiotic and synbiotic supplementation on ponderal and linear growth in severely malnourished young infants in a randomized clinical trial. Sci Rep. 2023;13(1):1845.

Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, et al. Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function. 2012;336(6080):489-93.

Pluske JR, Turpin DL, Kim JC. Gastrointestinal tract (gut) health in the young pig. Anim Nutr. 2018;4(2):187-96.

Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS One. 2020;15(1):e0227434.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat. biotechnol. 2017;35(9):833-44.

Rackaityte E, Halkias J, Fukui EM, Mendoza VF, Hayzelden C, Crawford ED, et al. Viable bacterial colonization is highly limited in the human intestine in utero. Nat Med. 2020;26(4):599-607.

Radhakrishnan ST, Gallagher KI, Mullish BH, Serrano-Contreras JI, Alexander JL, Miguens Blanco J, Danckert NP, Valdivia-Garcia M, Hopkins BJ, Ghai A, Ayub A. Rectal swabs as a viable alternative to faecal sampling for the analysis of gut microbiota functionality and composition. Sci. Rep. 2023 ;13(1):493.

Rai SN, Qian C, Pan J, Rai JP, Song M, Bagaitkar J, et al. Microbiome data analysis with applications to pre-clinical studies using QIIME2: Statistical considerations. Genes Dis. 2021;8(2):215-23.

Raman AS, Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Subramanian S, et al. A sparse covarying unit that describes healthy and impaired human gut microbiota development. Science. 2019;365(6449).

Ravisankar S, Tatum R, Garg PM, Herco M, Shekhawat PS, Chen YH. Necrotizing enterocolitis leads to disruption of tight junctions and increase in gut permeability in a mouse model. BMC Pediatr. 2018;18(1):372.

Razali H, O'Connor E, Drews A, Burke T, Westerdahl H. A quantitative and qualitative comparison of illumina MiSeq and 454 amplicon sequencing for genotyping the highly polymorphic major histocompatibility complex (MHC) in a non-model species. BMC Res Notes. 2017;10(1):346.

Reeder J, Knight R. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Methods. 2010;7(9):668-9.

Reyman M, van Houten MA, van Baarle D, Bosch AA, Man WH, Chu ML, Arp K, Watson RL, Sanders EA, Fuentes S, Bogaert D. Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. Nat. commun. 2019 ;10(1):4997.

Robertson RC, Manges AR, Finlay BB, Prendergast AJJTim. The human microbiome and child growth– first 1000 days and beyond. 2019;27(2):131-47.

Rosa F, Michelotti TC, St-Pierre B, Trevisi E, Osorio JS. Early life fecal microbiota transplantation in neonatal dairy calves promotes growth performance and alleviates inflammation and oxidative stress during weaning. Animals. 2021 ;11(9):2704.

Rosen MJ, Callahan BJ, Fisher DS, Holmes SP. Denoising PCR-amplified metagenome data. BMC bioinformatics. 2012 ;13(1):1-6.

Rothberg JM, Leamon JH. The development and impact of 454 sequencing. Nat Biotechnol. 2008;26(10):1117-24.

Saffarian A, Mulet C, Regnault B, Amiot A, Tran-Van-Nhieu J, Ravel J, Sobhani I, Sansonetti PJ, Pédron T. Crypt-and mucosa-associated core microbiotas in humans and their alteration in colon cancer patients. MBio. 2019 ;10(4):e01315-19.

Salgar-Chaparro SJ, Machuca LL. Complementary DNA/RNA-Based Profiling: Characterization of Corrosive Microbial Communities and Their Functional Profiles in an Oil Production Facility. Front Microbiol. 2019;10:2587.

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC biology. 2014 ;12:1-2.

Sangild PT, Strunk T, Currie AJ, Nguyen DN. Editorial: Immunity in Compromised Newborns. Front Immunol. 2021;12:732332.

Satokari R, Gronroos T, Laitinen K, Salminen S, Isolauri E. Bifidobacterium and Lactobacillus DNA in the human placenta. Lett Appl Microbiol. 2009;48(1):8-12.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41.

Schokker D, Zhang J, Zhang LL, Vastenhouw SA, Heilig HG, Smidt H, et al. Early-life environmental variation affects intestinal microbiota and immune development in new-born piglets. PLoS One. 2014;9(6):e100040.

Seitz V, Schaper S, Droge A, Lenze D, Hummel M, Hennig S. A new method to prevent carry-over contaminations in two-step PCR NGS library preparations. Nucleic Acids Res. 2015;43(20):e135.

Siegwald L, Caboche S, Even G, Viscogliosi E, Audebert C, Chabe M. The Impact of Bioinformatics Pipelines on Microbiota Studies: Does the Analytical "Microscope" Affect the Biological Interpretation? Microorganisms. 2019;7(10).

Slanzon GS, Ridenhour BJ, Parrish LM, Trombetta SC, Moore DA, Sischo WM, McConnel CS. Effects of a farm-specific fecal microbial transplant (FMT) product on clinical outcomes and fecal microbiome composition in preweaned dairy calves. Plos one. 2022 ;17(10):e0276638.

Song SJ, Wang J, Martino C, Jiang L, Thompson WK, Shenhav L, et al. Naturalization of the microbiota developmental trajectory of Cesarean-born neonates after vaginal seeding. Med (N Y). 2021;2(8):951-64 e5.

Stevenson DM, Weimer PJ. Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Appl Microbiol Biotechnol. 2007;75(1):165-74.

Stinson L, Hallingstrom M, Barman M, Viklund F, Keelan J, Kacerovsky M, et al. Comparison of Bacterial DNA Profiles in Mid-Trimester Amniotic Fluid Samples From Preterm and Term Deliveries. Front Microbiol. 2020;11:415.

Strathdee F, Free A. Denaturing gradient gel electrophoresis (DGGE). DNA electrophoresis: Methods Protoc. 2013:145-57.

Swanson KS, Gibson GR, Hutkins R, Reimer RA, Reid G, Verbeke K, et al. The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of synbiotics. Nat Rev Gastroenterol Hepatol. 2020;17(11):687-701.

Telatin A. Qiime Artifact EXtractor (Qax): A Fast and Versatile Tool to Interact with QIIME2 Archives. BioTech. 2021 ;10(1):5.

Thomson CA, Morgan SC, Ohland C, McCoy KD. From germ-free to wild: modulating microbiome complexity to understand mucosal immunology. Mucosal Immunol. 2022;15(6):1085-94.

Vogt SL, Finlay BB. Gut microbiota-mediated protection against diarrheal infections. J Travel Med. 2017;24(suppl_1):S39-S43.

Wang J, Tian S, Yu H, Wang J, Zhu W. Response of Colonic Mucosa-Associated Microbiota Composition, Mucosal Immune Homeostasis, and Barrier Function to Early Life Galactooligosaccharides Intervention in Suckling Piglets. J Agric Food Chem. 2019;67(2):578-88.

Wang L, Cao H, Liu L, Wang B, Walker WA, Acra SA, et al. Activation of epidermal growth factor receptor mediates mucin production stimulated by p40, a Lactobacillus rhamnosus GG-derived protein. 2014;289(29):20234-44.

Wang X, Jordan IK, Mayer LW. A phylogenetic perspective on molecular epidemiology. Mol. microbiol. 2015 Jan 1:517-36.

Wang Y, Gong L, Wu YP, Cui ZW, Wang YQ, Huang Y, et al. Oral administration of Lactobacillus rhamnosus GG to newborn piglets augments gut barrier function in pre-weaning piglets. J Zhejiang Univ Sci B. 2019;20(2):180-92.

Weese JS, Jelinski M. Assessment of the Fecal Microbiota in Beef Calves. J Vet Intern Med. 2017;31(1):176-85.

Weyrich LS, Duchene S, Soubrier J, Arriola L, Llamas B, Breen J, et al. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. Nature. 2017;544(7650):357-61.

Wu Y, Nie C, Luo R, Qi F, Bai X, Chen H, Niu J, Chen C, Zhang W. Effects of multispecies probiotic on intestinal microbiota and mucosal barrier function of neonatal calves infected with E. coli K99. Front. Microbiol. 2022:3933.

Xiang Q, Yan X, Shi W, Li H, Zhou K. Early gut microbiota intervention in premature infants: Application perspectives. J Adv Res. 2022.

Yamamoto, Masahiro, Rui Yamaguchi, Kaori Munakata, Kiyoe Takashima, Mitsue Nishiyama, Kyoji Hioki, Yasuyuki Ohnishi et al. A microarray analysis of gnotobiotic mice indicating that microbial exposure during the neonatal period plays an essential role in immune system development. BMC genomics 13, no. 1 (2012): 1-15.

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486(7402):222-7.

Zhang X, Zhivaki D, Lo-Man R. Unique aspects of the perinatal immune system. Nat Rev Immunol. 2017;17(8):495-507.

Zheng W, Tsompana M, Ruscitto A, Sharma A, Genco R, Sun Y, et al. An accurate and efficient experimental approach for characterization of the complex oral microbiota. Microbiome. 2015;3:48.

Zhou MI, Hernandez-Sanabria E, Guan LL. Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. AEM. 2009 ;75(20):6524-33.

Zhou P, Chen R, Zou CY, Zhang M, Chen ZH, Ma LY. Effect of perinatal factors on postnatal development of lymphocyte subsets in early preterm infants. Chin. J. Contemp. Pediatr. 2016 ;18(10):953-9.

Zhu L, Luo F, Hu W, Han Y, Wang Y, Zheng H, et al. Bacterial Communities in the Womb During Healthy Pregnancy. Front Microbiol. 2018;9:2163.

Zitvogel L, Kroemer G. Commensals shape the immune system. Nat. Rev. Immunol. 2021 ;21(10):615-.