Exploration and comparison of bacterial communities present in bovine milk, faeces and blood using 16S rRNA metagenomic sequencing

K Mtshali

orcid.org/0000-0003-3493-4066

Thesis accepted in fulfilment of the requirements for the degree *Doctor of Philosophy in Science with Zoology* at the North-West University

Promoter: Prof MMO Thekisoe

Co-promoter: Dr ZTH Khumalo

Graduation July 2022

28082419

DECLARATION

I, the undersigned, hereby declare that the work contained in this dissertation is my original work and that I have not previously in its entirety or in part submitted at any university for a degree. I furthermore cede copyright of the dissertation in favour of the North−West University.

Signature:  .....

Date: 17/ 06/ 2022 .....

DEDICATION

My Lord and my God!

‘For I know the plans I have for you, declares the Lord, plans for welfare and not for evil, to give you a future and a hope’ - Jeremiah 29:11

I dedicate this thesis to my deceased parents who brought me into this world, Mrs Jabulile Mtshali and Mr Robert Mtshali, may their souls rest in peace (RIP) and perpetual light shine upon them. I am also thankful to all my siblings, particularly Hlengiwe, Lindiwe and Bongiwe for the undying love and support throughout this journey. A special sense of gratitude to my psychospiritual team, close and distant: Reverend (Rev) Father (Fr) Angelicus Mchunu, Rev Fr Finbarr Flanagan, Rev Fr Dr John Selemela, Rev Fr Joel Skhosana, Rev Fr Thabiso Ledwaba, Rev Fr Xolani Ndaba, Rev Fr Johannes Sibanyoni, Rev Fr Thembelani Ngcobo, Rev Sister (Sr) Dominica Shezi, Rev Sr Joseph Ngcobo (RIP), Rev Sr Patience Mhlongo (RIP).

ACKNOWLEDGEMENTS

In this section, the following organizations and people that contributed to the success of this study in one way or another are acknowledged.

First and foremost, the National Research Foundation – Thuthuka PhD Track Funding Instrument (Unique Grant No. 107416) in collaboration with Tshwane University of Technology (TUT) for co - funding the study.

Bursaries granted to the principal investigator (PI) including TUT Study Grant for Permanent Staff (External Studies), North West University (NWU) Merit Bursary, NWU Doctoral Bursary, Faculty of Natural and Agricultural Sciences Postgraduate Bursary as well as tuition payment from the NRF Incentive Grant for Rated Researchers granted to Prof Oriel Thekisoe.

Provision of funding in kind and in the form of training workshops by the Technological Higher Education Network South Africa (THENSA) Staff Capacity Development PhD Program in collaboration with the Department of Higher Education and Training (DHET) and TUT.

The study promoters, Prof Oriel Thekisoe and Dr Zamantungwa Khumalo for their guidance in execution of the study. Dr Stanford Kwenda and Dr Moeti Taioe for assistance with data curation & software for data analysis.

For assistance with sampling and associated logistics, Mr James Morake (Control Animal Health Technician) and his team of Animal Health Technicians from uThukela Veterinary Services under the Ladysmith State Veterinary Animal Health division. A special sense of gratitude goes out to the animal owners and / or herders and the Waaihoek community dip site committee for allowing us to sample their animals for the purpose of this study.

Lastly, I am also thankful to Mr Marcus Makgabo from the Department of Veterinary Tropical Diseases (DVTD), University of Pretoria, Onderstepoort (UP-OP) who provided *Anaplasma* species positive controls as well as Prof Henriette van Heerden and Dr Betty Ledwaba from the same department who provided Biosafety Level 2 Plus (BSL2+) facility and *Brucella* positive controls, assisted with culturing, isolation and extraction of *Brucella* DNA.

RESEARCH OUTPUTS

Full-length research papers

Khethiwe Mtshali, Zamantungwa T.H., Khumalo, Stanford Kwenda, Ismail Arshad, Oriel M.M. Thekisoe. Exploration and comparison of bacterial communities present in bovine faeces, milk and blood using 16S rRNA metagenomic sequencing. PLOS One (currently under revision).

Khethiwe Mtshali, Zamantungwa T.H., Khumalo, Maphuti B., Ledwaba, Henriette van Heerden, Oriel M.M.,Thekisoe. 16S rRNA metagenomics as an important tool in detection of *Brucella* infection and characterization of associated taxa: A case report (draft paper).

Conference paper

Khethiwe Mtshali, Zamantungwa T.H., Khumalo, Stanford Kwenda, Ismail Arshad, Oriel M.M.,Thekisoe. Profiling Of Bacterial Communities Present In Feces, Milk, And Blood Of Lactating Cows Using 16S rRNA Metagenomic Sequencing. Oral presentation, ICVBM 2022: XVI. International Conference on Veterinary Biology and Microbiology, 16-17 May 2022, Amsterdam Netherlands.

Award

Best Presentation Award. Oral presentation, ICVBM 2022: XVI. International Conference on Veterinary Biology and Microbiology, 16-17 May 2022, Amsterdam Netherlands.

ABSTRACT

Livestock rearing plays a vital role in sustenance of the livelihoods of rural communities. In contrast, livestock serves as a potent reservoir of different pathogenic organisms that could have devastating health and economic implications, especially when proper husbandry and hygiene practices are not in place. Rural communities utilize cattle products and by-products to provide nourishment and income; to perform many cultural traditions and rituals; and to facilitate day-to-day household activities. Bovine faeces, milk and blood respectively derived from the gut, mammary glands and bloodstream niches are no exception and they are amongst the most commonly used by the Waaihoek community. Ecologically, these niches carry a diverse array of microbial communities of commensals, mutualists and pathogens. Although each of them harbours its own distinct and specialized microbial profile, there is an interplay of factors among these and various body niches which influence their colonization and assembly. Any imbalance in the structure of the microbes in this complex ecosystem of niches can lead to increased pathogenicity of constituent microbes and occurrence of diseases. This study aimed to simultaneously explore the microbiota of corresponding faecal, milk and blood samples from lactating cows using 16S rRNA metagenomic sequencing. A total of 24 sample pools were sequenced. Bacterial communities were inferred through the Divisive Amplicon Denoising Algorithm 2 (DADA2) pipeline coupled with SILVA database v138. All downstream analyses were performed in R v3.6.1. Alpha-diversity metrics showed significant differences between faeces and blood; faeces and milk; but did not vary significantly between blood and milk (Kruskal-Wallis, $P < 0.05$). Beta-diversity metrics on Principal Coordinate Analysis (PCoA) and Non-Metric Dimensional Scaling (NMDS) clustered samples by type suggesting that microbial communities of the studied niches are significantly different (PERMANOVA, $P < 0.05$). A number of taxa were significantly differentially abundant (DA) between groups based on the Wald test implemented in DESeq2 package (*Padj* < 0.01). Majority of the DA taxa (*i.e*. *Romboutsia*, *Paeniclostridium*, *Monoglobus*, *Akkermansia, Turicibacter, Bacteroides, Candidatus_Saccharimonas*, *UCG-005* and *Prevotellaceae_UCG-004*) were significantly enriched in faeces than in milk and blood, except for *Anaplasma* which was greatly enriched in blood and was in turn the most abundant taxon overall. The entire analysis revealed a total of 30 phyla, 74 classes, 156 orders, 243 families and 408 genera. A total of 58 genus-level taxa occurred concurrently between the niches, while bacterial signatures of at least 8 of these (*i.e*. *Romboutsia*, *UCG-005*, *Prevotellaceae_UCG-004*, *Rikenellaceae_RC9_gut_group*, *Bacteroides*, *Christensenellaceae_R-7_group*, *Turicibacter* and *Fusobacterium*) concurrently occurred in corresponding faeces, milk and

blood samples from the same group of animals constituting a pool. Firmicutes, Bacteroidota, and Proteobacteria were the most abundant phyla overall. The important taxa could be categorized into four pathogenic clusters: i) arthropod-borne; ii) food-borne and zoonotic; iii) mastitogenic and; iv) metritic and abortigenic. This study provides insight into the microbial composition of bovine faeces, milk, and blood and its extent of overlapping. It further highlights the potential risk of disease occurrence and transmission between the animals and the rural based communities pertaining to their unsanitary practices associated with the use of cattle by-products.

Key terms: Microbial profiling, 16S rRNA, NGS, faeces, milk, blood, lactating cows, small-scale farmers, Waaihoek, KwaZulu-Natal

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1. BACKGROUND OF THE STUDY

Metagenomics is a term coined around 1998 to describe a method that applies genome sequencing or assays of functional properties for the culture-independent analysis of complex and diverse ("meta") populations of microbes (Wooley & Ye 2009). It is loosely defined as the study of microbial communities, sampled directly from their natural environment without prior culturing (Wooley & Ye 2009; Shah *et al*., 2011; Kwong *et al*., 2015). The method either applied at large-scale (shotgun metagenomics) or small-scale (16S rRNA gene based or targeted) approaches, has carved a pathway into the study of microbial community structures, phylogenetic composition, species diversity, metabolic capacity, and functional diversity (Singh *et al*., 2009; Wooley & Ye 2009; Shah *et al*., 2011).

The development and evolution of omics technologies, the so-called Next Generation Sequencing (NGS) techniques from the early 2000s has greatly enhanced profiling of microbial communities and gained popularity among clinical diagnosticians (Wooley & Ye 2009). It has helped to improve knowledge and understanding of the course of infectious diseases and clinical microbiology (Kwong *et al.*, 2015), particularly in the public health sector.

The use of metagenomics has revolutionized the world view of microbial diversity (Gupta & Verma, 2019) with many niches which were traditionally thought to be sterile now being described as consisting of complex and diverse microbial communities of commensals, mutualists and some pathogenic agents (Castillo *et al*., 2019; Oikonomou *et al*., 2020). Steadily the veterinary sector has caught the wave of metagenomics and recently microbial investigations have been conducted on various body sites such as the gut, mammary glands, bloodstream as well as the oropharyngeal, respiratory, vaginal and uterine tracts (Jami *et al*., 2014; Jeon *et al*. 2017, Vidal *et al*., 2017; Vientós-Plotts *et al*., 2017; Oikonomou *et al*., 2020). The body sites form a complex of ecological niches in the mammalian body, harboring distinct and specialized microbial profiles (Derakhshani *et al*., 2018). There is an interplay of factors among various body sites which influence their colonization and microbial assembly (Derakhshani *et al*., 2018; Castillo *et al*., 2019). Any imbalance in the structure of the microbes in this complex ecosystem of niches can lead to increased pathogenicity of

constituent microbes and occurrence of diseases (Deng *et al*., 2019). In such instances, even commensals may suddenly act opportunistically to cause infections (Maity & Ambatipudi, 2021).

The bovine gut is one of the most studied ecological niches in ruminants using high-throughput NGS techniques. However, the majority of the research has been mainly focused on the rumen, linked to its important role in feed efficiency (FE) and less so on other components of the gastro-intestinal tract (GIT) (McSweeney and Mackie, 2012; O'Hara *et al*., 2020). Next Generation Sequencing based studies on the bovine mammary glands, milk and colostrum started to increase in the past decade (Oikonomou *et al*., 2012, 2013 & 2014; Addis *et al*., 2016; Falentin *et al*., 2016, Lima *et al*., 2017; Rainard, 2017; Derakhshani *et al*., 2018; Taponen *et al*., 2019; Oikonomou *et al*., 2020), however they have been mainly focused on how the microbial flora of milk changes when it becomes a food product, either for direct consumption or for transformation into dairy products (Addis *et al*., 2016). A few of these studies have looked into the microbiome of milk derived from mastic quarters (Oikonomou *et al*., 2012 & 2014; Falentin *et al*., 2016; Hoque *et al*., 2019 & 2020), but not much research has been conducted to determine the global diversity of milk microbiota in relation to udder health, disease and physiology (Addis *et al*., 2016; Derakhshani *et al*., 2018). On the contrary, NGS based studies on animal blood microbiomes are very rare. The few studies that have been conducted include characterization of the blood microbiome of cats, broilers and recently, cattle in relation to disease microbial dysbiosis (a term used to describe changes in a microbiome in its main origin, typically the gut) and atopobiosis (a term describing microbes that appear in places other than where they should be), (Potgieter *et al*., 2015; Mandal *et al*., 2016; Jeon *et al*., 2017; Vientós-Plotts *et al*., 2017). The bovine blood microbiome has been mainly probed in investigations of its role in translocation of microbes from the uterus or the gut to other body sites such as the mammary glands (Young *et al*., 2015; Galvao *et al*., 2017; Jeon *et al*., 2017).

According to Rodriguez *et al*., (2021) there is a complex and complementary interaction between the gut, mammary glands and the bloodstream which is quintessential for neonatal gut development. Furthermore, the authors state that the maternal gut supplies the mammary glands with an array of nutrients & bioactive components, including microbes, immune cells and stem cells, which are then passed through the milk to the offspring and help with development of immune tolerance and life-long imprinting of the immune system. The translocation of microbes in particular, is said to occur via an endogenous entero-mammary pathway where they are trafficked from the gut via the bloodstream to the mammary glands (Rodriguez 2014; Rodriguez *et al*., 2021). The pathway is largely studied in humans and other monogastric animals, but its existence is highly disputed in ruminants as the link

between the immune system of the udder and that of the intestine is said to be very poor in these animals (Derakhshani *et al*., 2018; Rainard, 2017). Although not providing compelling evidence of this occurrence, a recent investigation of the link between the three niches in lactating cows has demonstrated simultaneous occurrence of identical bacterial signatures from faeces, milk somatic cells and bloodstream macrophages (Young *et al*., 2015).

Understanding the structure and composition of microbiota in the gut, mammary glands and the bloodstream regardless of whether their presence is beneficial, inconsequential or detrimental, is critical in understanding disease pathogenesis, diagnosis and therapeutic intervention as well as improvement of health and livestock productivity.

1.2. PROBLEM STATEMENT

South African agriculture is comprised of mainly two categories of farmers *i.e.* small-scale subsistence and large-scale commercial farmers. This study focused on the former, which in the South African context is defined as one whose scale of operation is too small to attract the provision of the services he/she needs to be able to significantly increase his/her productivity (Kirsten & Van Zyl, 1998).

Subsistence farming plays a vital role in sustenance of the livelihoods of rural communities (Ducrotoy *et al*., 2015). However, livestock serve as a potent reservoir of different pathogenic organisms that could have devastating health and economic implications, especially when proper husbandry and hygiene practices are not in place (Marufatuzzahan *et al*., 2018). It may additionally have compounding effects on the public health due to the zoonotic nature of some of the associated pathogenic agents (Kaoud *et al*., 2019).

Primarily, cattle products and by-products derived from these farms are utilized to provide nourishment and income; secondarily, they play important social roles in many cultural traditions and rituals; and help to facilitate day-to-day household activities (McDermott and Arimi, 2002;

Marufatuzzahan *et al*., 2018). Bovine faeces, milk, and blood are no exception and they are among the most commonly used cattle by-products particularly by the targeted rural community.

South Africa, like many other developing countries globally, comprises the largest traditional livestock farming system. The farmers are mostly undereducated with no proper training and possess less knowledge about animal health, disease control, animal development issues as well as policies regarding animal production (DAFF, 2018; Marufatuzzahan *et al*., 2018). Majority of them manage their farms the old-traditional ways, without any concept or concern of modern and healthy farm management practices. Sometimes, these farmers have the facilities but do not follow proper hygiene as they are not well-educated about the consequences of unhealthy farming systems (Marufatuzzahan *et al*., 2018). They mostly rear their livestock in close proximity to households which poses a major risk of transmission of pathogens between animals and humans as has been previously expounded by various authors (Tschopp *et al*., 2013; Ducrotoy *et al*., 2015; Lorusso *et al*., 2016; Maity & Ambatipudi, 2021). Moreover, the animals are packed in high stocking densities within the enclosures which also increases the risk of transmission of pathogens between the animals (Kaoud, 2019).

Thus, with or without knowledge, the above-mentioned 'unsanitary' practices and the incorrect and backward farming practices implemented by small-scale subsistence farmers contribute to the occurrence and spread of livestock associated diseases. This lead to the formulation of the first research question: i) What constitutes the microbial communities of faeces, milk and blood of cattle owned by the small-scale subsistence farmers? Case in point, the Waaihoek community.

The second research question was formulated based on the current knowledge of the complex and complementary interaction of the gut, bloodstream and mammary gland microbiomes in monogastric animals (Rodriguez 2014, Rodriguez *et al*., 2021); the hypothesized endogenous route of translocation of microbes from the bovine gut to the mammary glands via the bloodstream (Quigley *et al*., 2013; Young *et al*., 2015); and from the previous concurrent detection of identical bacterial signatures in bovine faeces, milk somatic cells and bloodstream macrophages (Young *et al*., 2015). It reads as follows: ii) Are there any similarities in the microbial communities contained in corresponding samples of bovine faeces, milk and blood?

Over the years, culture-dependent techniques have proven problematic for diagnosis of a significant fraction of Bacteria (Klindworth *et al*., 2013), with over 99.8% of the microbes in some niches remaining uncultivable in growth media (Singh *et al*., 2009; Shah *et al*., 2011). Therefore, the majority

of what is known about microbial diversity & function in different niches is biasedly based on the small fraction (<1%) of cultivable species. This lead to the formulation of the third and last research question: iii) How can the structure and composition of microbial communities contained in bovine faeces, milk and blood be explored without the limitations presented by culture-dependent techniques?

Thus in an attempt to answer the posed research questions, the structure and composition of microbial communities present in faeces, milk and blood of lactating cows, respectively representing the gut, mammary glands and the bloodstream were explored through culture-independent metagenomic sequencing. This was achieved through amplicon targeted 16S rRNA sequencing of the V3 - V4 hypervariable region at the employ of the Illumina Miseq protocol with anticipation to identify much greater diversity than standard culture methods. The choice of region particularly influenced by its reported accuracy in differentiation of bacterial genera and its frequent use in surveys of samples from various mammalian hosts (Shah *et al*., 2011; Young *et al*., 2015; Rausch *et al*. 2019; Oikonomou *et al*., 2020).

1.3. AIMS AND OBJECTIVES OF THE STUDY

1.3.1. Aims

Primarily, the aim of the study was to explore the microbial community structure and composition of corresponding faeces, milk and blood of lactating cows owned by small-scale subsistence farmers in a rural community using 16S rRNA metagenomic sequencing. Secondarily, to identify the shared and differentially abundant (DA) taxa between the three niches. Lastly, to identify and characterize selected pathogens of veterinary and / or medical significance yielded through the primary and secondary aims of the study.

1.3.2. Objectives

i. To characterize the taxonomic profiles of microbes present in bovine faeces, milk and blood using 16S rRNA metagenomic sequencing.

ii. To determine the microbial community structure of bovine faeces, milk and blood samples using alpha- and beta- diversity metrics.
iii. To determine and compare shared and differentially abundant microbial taxa between bovine faeces, milk and blood.
iv. To determine the prevalence of taxa of veterinary and / or medical importance contained within bovine faeces, milk and blood.
v. To characterize selected taxa of veterinary and / or medical significance obtained through 16S rRNA metagenomic sequencing using species specific Polymerase Chain Reaction (PCR).

1.4. THESIS OUTLINE

CHAPTER 1: INTRODUCTION

This section provides the background and problem statement in efforts to justify the significance of this study. The research questions, aims and the objectives of the study are also addressed here.

CHAPTER 2: LITERATURE REVIEW

This chapter gives the literature review to the aims and objectives of the study where the microbial composition of faeces, milk and blood, as well as the tools used to explore these microbiomes and their efficacy are discussed.

CHAPTER 3: MATERIALS AND METHODS

This chapter covers the research design and details the methodology used in the current study. Briefly, the methods of sample collection, DNA extraction, quality filtering and quantification are discussed. Amplification and sequencing of the V3-V4 hypervariable region of the bacterial 16S rRNA gene on the Illumina MiSeq platform is outlined. Extraction of paired-end sequences from the Illumina platform, quality inspection, trimming, dereplication, merging of paired-end reads and removal of chimeric sequences as well as bioinformatics tools used to plot and analyze the sequencing data

obtained in the study are outlined. The methods of detection and characterization of selected pathogenic species of veterinary and / or medical significance using PCR are outlined.

CHAPTER 4: RESULTS

The microbial communities contained within bovine faeces, milk and blood are characterized, enumerated and their abundances and prevalence determined within and between samples. Followed by determination of the shared and differentially abundan taxa between bovine faeces, milk and blood.

The structure (α- and β-diversity) of communities hosted within faeces, milk, blood and coincidentally placenta is determined and a comparative analysis of the microbial communities hosted within and between the sample groups is conducted using various analytical statistical tools applied in microbial ecology and the results thereof analyzed.

Results on the incidence and PCR characterization of pathogens of veterinary and medical significance are analyzed.

CHAPTER 5: DISCUSSION, CONCLUSION AND RECOMMENDATIONS

The overall findings of the study are discussed including the potential health implications of the findings for the sampled rural small-scale farming community. Conclusions are drawn and recommendations are made on the limitations and value of the study's findings, currently and for future research.

CHAPTER 2: LITERATURE REVIEW

2.1. THE MICROBIOME

Microorganisms are the first life forms, dating back 3.5 billion years ago according to fossil evidence. The current estimate of the number of prokaryotic cells on earth is around 4×10^{30} - 6×10^{30}, comprising 10^6 and 10^8 separate genospecies Singh *et al*., (2009). According to these authors, the microorganisms represent an enormous but largely unexplored genetic and biological pool that can be exploited for the recovery of novel genes, entire metabolic pathways and their products.

Microorganisms live in complex communities of interacting species that impact life on earth and geochemical processes in the environment (Milanese *et al*., 2019). It is thus of fundamental interest to accurately profile and compare the composition of the communities which they form.

The mammalian host harbours a complex ecosystem of bacteria, fungi, protozoa and viruses, referred to as the microbiome. It begins to form prior to birth, in the uterus, developing for a few years after birth to become a stable, fully functioning microbiome, until the physiological changes associated with senescence lead to substantial shifts in its composition (Alipour *et al*., 2018; Klimesova *et al*., 2018). Colonization of the gut typically occurs *per os*, moreover, via the bloodstream the microbiome can disseminate all over the body (Klimesova *et al*., 2018).

This process of colonization is influenced by the interaction between the microbes, host, and environment (Malmuthuge *et al.,* 2015). Microbial factors to name a few, include adhesion, survival mechanisms under oxygen gradient, and mechanisms to obtain nutrients from the host; host factors may include luminal pH, food retention time in the gut and immune defense mechanisms; while external factors include but are not limited to the maternal microbiota, delivery mode, diet, and antibiotic treatment during early life as can be seen on Fig 2.1 (Malmuthuge *et al.,* 2015; Klimesova *et al*., 2018). These factors combine to influence the composition of the microbiome of the host (Malmuthuge *et al.,* 2015; Klimesova *et al*., 2018; Zeineldin *et al*., 2018).

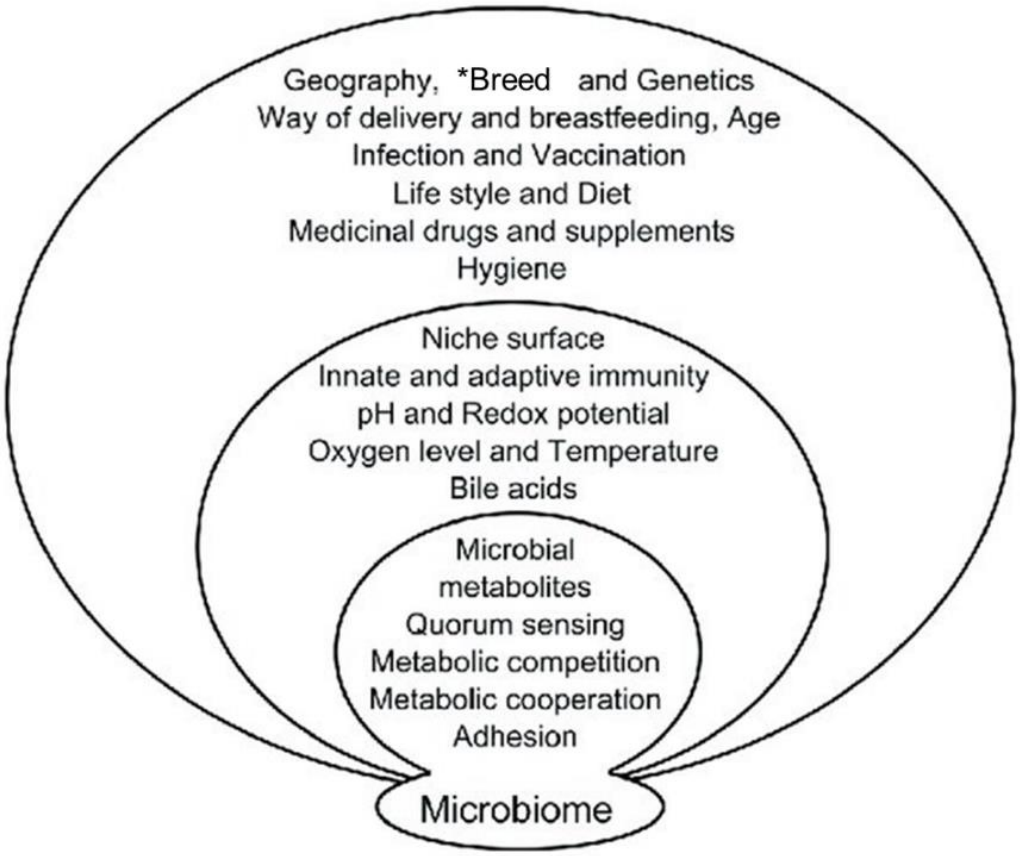

Fig 2.1: Factors influencing microbiome composition of a host. The asterisk (*) indicates a change in the original diagram where the word 'Ethnicity' was replaced with 'Breed'. Diagram adapted from Klimesova *et al*., (2018).

In the veterinary sector, identification, characterization and enumeration of the microbial organisms with pathogenic potential in clinical samples is an important component of surveillance and eventual clinical resolution of various diseases (Rapp, 2010). This is usually done with the intent of giving insight into the etiological agent(s) causing an infectious disease, including pathological associations and possible effective antimicrobial therapy (Clarridge, 2004).

Microbial ecology has developed as a specialized research field in microbiology and focuses on studying the role of microbes in a variety of ecosystems in the mammalian body which include body tissues and biofluids such as the blood, bone marrow, gut, milk and the trachea (Mandal *et al*., 2016). The main focus among animal microbiologists and ruminant health scientists has been on studying the composition of the microbiota of three organs *i.e.* rumen, udder, and uterus in dairy cows (Dias

and Ametaj, 2017). For the purpose of this study the focus was on reviewing the microbiota of the gut, milk, and blood.

2.1.1. Gut microbiota (GM)

The microbial ecology of faeces cannot be studied as a separate entity as it is closely linked to the microbiome of the gut (in this context referring to the GIT which begins from the oral cavity through to the rectum). It has been established that the GM diverges in composition and function according to region throughout the gut, echoing the differences in physical, chemical and biological conditions in each compartment (Klimesova *et al*., 2018; O'Hara *et al*., 2020). Because faeces is a by-product of all the metabolic processes (including digestion and absorption) that occur throughout the gut after ingestion, its microbial composition within the rectum is thus shaped by the GM.

According to McSweeney & Mackie, (2012), the study of the microbial ecology of the gut begins with investigation of the organisms present (abundance and diversity) and subsequently succeeded by investigation of their activity and relationships with each other and the host animal (synergistic and competitive interactions).

The *in utero* sterile mammalian gut is said to be rapidly colonized by an array of microbiota before, during and after birth (Malmuthuge *et al.,* 2015). Major colonization however starts at birth and is complemented during lactation and later life (Alipour *et al*., 2018). This colonization is influenced by the microbial, host and external factors as discussed above (Malmuthuge *et al.,* 2015; Klimesova *et al*., 2018).

The initial colonizers (*e.g. Streptococcus* and *Enterococcus* species; spp.) utilize available oxygen in the gut and create the anaerobic environment required for strict anaerobic gut residents, such as *Bifidobacterium* and *Bacteroides*. These microbes are two of the main gut bacteria that have a beneficial impact on the mucosal immune system where they play a vital role in the development of immunological tolerance to commensal microbiota (Malmuthuge *et al.,* 2015, O'Hara *et al*., 2020). Therefore, neonatal gut colonization is a crucial period for the developing gut and naive immune system and may have long-term health effects (Malmuthuge *et al.,* 2015).

Although research focused on understanding gut colonization of mammals has increased dramatically over the years, there are still very few studies focused on domestic livestock species,

especially ruminants (Dowd *et al*., 2008; Malmuthuge *et al.,* 2015; Dias and Ametaj, 2017). The dramatic increase can be attributed partly to the fact that numerous human gut infections are being linked to the consumption of meat contaminated with faeces, from gut content during the slaughter process. Other studies have identified as a risk factor the consumption of lettuce, spinach and other crops which are fertilized using animal manure (Robinson *et al*., 2009; Rapp, 2010; McSweeney& Mackie, 2012). Furthermore, the burgeoning of novel tools (primarily NGS) in the markets within the past decade and a half has enabled the assessment of genes and genomes contained within complex microbial communities such as faeces and contributed to understanding of many human diseases (Panek *et al*., 2018), and to some extent animal diseases.

Most of the newly sprouted research on the GM of ruminants has been mainly focused on the microbiome of the rumen, in contrast there has been little research on other parts of the bovine gut. For the purpose of this study the microbial composition of the gut will be reviewed under three subtopics *i.e.* the rumen, lower gut and faeces. From the existing literature on investigations of the GM and gut colonization, there is an observed difference in the microbiome of the pre-ruminant versus ruminant gut. There are a variety of factors which greatly influence its establishment from neonatal development throughout the life of the animal and these are discussed under the subtopics.

2.1.1.1. Microbial composition of the rumen

The rumen microbiome consists of high densities of bacteria (10^{10}-10^{11} g^{-1}; > 200 species), archaea (10^{6}-10^{8} g^{-1}), protozoa (10^{4}-10^{6} g^{-1}, 25 genera), bacteriophages (10^{4}-10^{6} g^{-1}, 25 genera) and fungi (10^{3}-10^{5} g^{-1}, 5 genera) involved in the fermentation of complex carbohydrates. The composition of these microbes is influenced by a number of factors which are dependent on the ruminant species, host age, diet, season and geographic region (McSweeney & Mackie, (2012); Malmuthuge *et al.,* 2015; Zeineldin *et al*., 2018; O'Hara *et al*., 2020).

The microbial cohort contained within the rumen include cellulolytic, hemicellulolytic, amylolytic, proteolytic and biohydrogenating (lipolytic) species, exhibiting a high level of functional redundancy and capable of effectively degrading host-indigestible plant fiber (McSweeney& Mackie, 2012; O'Hara *et al*., 2020). Bacteria dominate the rumen microbiome and contribute mainly to the production of volatile fatty acids (VFAs; principally, acetate and propionate) as well as production of microbial protein. These are subsequently absorbed and used as energy sources by the host (McSweeney& Mackie, 2012; Malmuthuge *et al*, 2015; O'Hara *et al*., 2020).

The microorganisms provide 70% of the daily energy requirements of the host through the fermentation of indigestible dietary substrates, thus playing an essential role in the development of ruminants (McSweeney & Mackie, 2012; Ross *et al*., 2012**;** Malmuthuge *et al.,* 2015; Oyama *et al*., 2017; O'Hara *et al*., 2020). Moreover, the microbial composition of the rumen is associated with variations in feed efficiency (FE), intensity of methane gas (CH_4) emission, health status of the animal and milk composition (Ross *et al*., 2012; O'Hara *et al*., 2020), thus perpetuating more research based on this region of the gut.

Early studies of the pre-ruminant rumen colonization date back to the 1940s, studied using light microscopy coupled with Gram staining to visualize the bacteria. The most prominent early rumen microbiota by bacterial culture consisted of bacterial species from the genera *Propionibacterium*, *Clostridium*, *Peptostreptococcus* and *Bifidobacterium*; while *Ruminococcus* species are said to dominate the cellulolytic bacterial population. The changes in the appearance of the neonatal rumen microbiota are reportedly age-dependent (Malmuthuge *et al.,* 2015).

NGS-based studies have revealed a taxonomically and functionally diverse microbiome in pre-ruminant calves with significant age-dependent changes (Malmuthuge *et al.,* 2015), similar to culture-based observation as stated above. Members of Bacteroidetes (now Bacteroidota), followed by Firmicutes and Proteobacteria have been found to be among the initial phyla to colonize the ruminal content of pre-weaned calves (Malmuthuge *et al.,* 2015).

The microbial cohort found in the ruminating gut is established during the pre-ruminant calf stages and evolves over time due to the above-mentioned factors. The most numerous microbial group in the rumen of ruminating animals are bacteria, dominated by members of the Firmicutes, Bacteroidota and Proteobacteria phyla, containing numerous genera like *Prevotella*, *Fibrobacter* and *Butyrivibrio* which are capable of metabolizing a range of dietary polysaccharides and peptides (McSweeney& Mackie, 2012; Zeineldin *et al*., 2018; O'Hara *et al*., 2020). More than 90% of the Firmicutes sequences have been assigned to genera within the class Clostridia while streptococci are prominent within the class Bacilli. Within Clostridia; Lachnospiraceae, Ruminococcaceae and Veillonellaceae are the largest families reported. The predominant genera within this phylum are said to include *Butyrivibrio, Acetivibrio*, *Ruminococcus*, *Succiniclasticum*, *Pseudobutyrivibrio* and *Mogibacterium* (McSweeney& Mackie, 2012; Zeineldin *et al*., 2018).

It is a known fact that milk bypasses the rumen to enter the abomasum of the suckling calf, thus the establishment of the rumen microbiome is largely dependent on the introduction of a solid diet (Porter, 1969). Furthermore, pre-weaning diet and feeding methods have been reported to have more pronounced and long-lasting impacts on rumen microbial composition (Malmuthuge *et al.,* 2015). Different studies have reported different findings regarding the abundance of early colonizers of the rumen and it is apparent that among the many factors which may influence the rumen microbiome, diet, age and environmental exposure take the lead.

2.1.1.2. Microbial composition of the lower gut

According to O'Hara *et al*., (2020), the lower gut is defined as the post-gastric intestinal tract, consisting of both the small intestine and the hindgut region including the colon, caecum and rectum. Thus, the microbiota from these regions will hereinafter be referred to as the lower gut microbiota (GM). In contrast to that of the rumen, the fundamental role(s) of the lower GM and its contribution to ruminant health and production are poorly understood.

Characteristically, the lower GM diverges in composition according to intestinal segment (O'Hara *et al*., 2020). However, investigations on the composition of intestinal microbiota in ruminants are incredibly scarce and warrant further research (Dias & Ametaj, 2017).

Early investigation of bacterial colonization in the pre-ruminant lower gut dates back to 1965. The researchers used culture-dependent approaches to study first time colonizers immediately postpartum. Similarly to the rumen, the first time colonizers of this region are facultative anaerobes, which create the conditions required for colonization by obligate anaerobic GM such as *Lactobacillus* and *Bacteroides*. *Escherichia coli* and *Streptococcus* spp. are found to be among the predominant colonizers in the lower gut, with increase in *Lactobacillus*, *Bacteroides* and *Clostridium perfringens* in the caecum and faeces in the first week of life (Malmuthuge *et al.*, 2015).

Similar findings have been recorded using culture-independent approaches in newborn calves where there were higher abundances of *Bifidobacterium* and *Lactobacillus* in the duodenum and colon (Malmuthuge *et al.,* 2015). Large intestinal regions consisted primarily of Bacteroidota and Firmicutes, while >95% of the bacteria in the small intestinal contents were composed of Firmicutes (Malmuthuge *et al.,* 2015). The neonatal rectal microbiota is reportedly composed of Firmicutes, Proteobacteria, Actinobacteria (now Actinobacteriota) and Bacteroidota with *Escherichia/Shigella*

and Clostridia as the predominant taxa (Alipour *et al*., 2018). This neonatal rectal microbial profile closely resembled that of the dam's oral cavity, rather than the faecal or vaginal vestibular microbiota, particularly highlighting the role of the diet of the pregnant female in seeding the gut of the developing foetus (Alipour *et al*., 2018).

The ruminating lower gut microbiome also varies in proportions of microbes per segment with significant age-dependent changes in its composition, furthermore it is substantially different from that of the rumen and the pre-ruminating lower gut (Malmuthuge *et al.,* 2015).

The jejunum is mainly colonized by Firmicutes (between 90 - 95%) as the predominant phyla (McSweeney & Mackie, 2012; O'Hara *et al*., 2020). The hindgut regions are mainly predominated by microbial communities belonging to Firmicutes and Bacteroidota phyla (Dias & Ametaj, 2017; O'Hara *et al*., 2020). Several taxa in both the small and large intestine have been related to FE status, with divergent abundances across FE phenotypes *i.e. Butyrivibrio, Pseudobutyrivibrio, Prevotella*, *Anaeroplasma*, *Paludibacter*, *Faecalibacterium* and *Succinivibrio* in the hindgut, and mainly *Butyrivibrio* in the jejunum. The microbial communities of the lower gut indicate that it may indeed be closely related to cattle production efficiency (O'Hara *et al*., 2020).

Bacteria are present at levels of 10^{12}-10^{14} g^{-1} in the hindgut digesta of ruminating cattle. Microbial fermentation in the hindgut may be responsible for up to 30% of cellulose and hemicellulose degradation in ruminants. This lower dietary energy production in the hindgut compartments is likely due to a combination of factors including reduced retention time of digesta when compared to the rumen, as well as the fact that substrates entering the cecum and colon already have been partially digested by enzymes in the rumen (microbial) and small intestine (host and microbial), (O'Hara *et al*., 2020).

In addition to their presumed role(s) in feed digestion and energy production, there is also increasing evidence that the microbial communities of the lower gut contribute to immune system establishment and homeostasis in beef cattle that directly impact animal gut health (O'Hara *et al*., 2020).

2.1.1.3. Microbial composition of faeces

Characteristically, bovine faeces consists of a large number of endogenous (from the animal) and exogenous (from the diet) components which have passed through the animal's gut. These

components include diverse microorganisms and a high (47 - 68%) fibre content because of undigested residues of herbivorous matter *i.e.* cellulose, hemicelluloses, and lignin (Rapp, 2010).

Ultimately, the metabolic processes occurring in the different regions largely determine the abundance and diversity of the various communities of species present in the faeces, with an overlap of about 45% microbial species found in the oral cavity being present in faeces of tested individuals (Klimesova *et al*., 2018). The diverse and numerous gut microbial populations often exceed 10^{11} colony forming units per gram (CFU/g) of faeces (Dowd *et al*., 2008).

No faecal deposit is similar to another in composition. This has been proven through investigations of heterogeneity in distribution of microbial pathogen cells within faecal deposits (Pearce *et al*., 2004; Robinson *et al*. 2005; Rapp, 2010). Once in the environment there is an additional component of contamination, thus signifying the essence of sterile sampling techniques, *i.e.* collection of faeces directly from the rectum, when conducting microbial profiling studies.

High-throughput sequencing (HTS) of bovine faecal microbes often reveals the presence of the Firmicutes, Bacteroidota, Tenericutes, Cyanobacteria, Proteobacteria and Actinobacteriota (Young *et al*., 2015; Dias & Ametaj, 2017; Koester *et al*., 2020). The predominating taxa include numerous genera such as *Ruminococcus, Clostridium, Porphyromonas, Bacteroides, Alistipes, Lachnospira* and *Prevotella*; various unclassified members of the order Bacteroidales and unclassified taxa in the families *Ruminococcaceae*, *Bacteroidaceae*, *Lachnospiraceae*, *Paraprevotellaceae* and *Rikenellaceae* (Dowd *et al*., 2008; Young *et al*., 2015).

The predominance of Firmicutes, Bacteroidota and Proteobacteria is seen throughout the gut (Fig.2.2) and ultimately in faeces (Dias and Ametaj, 2017).

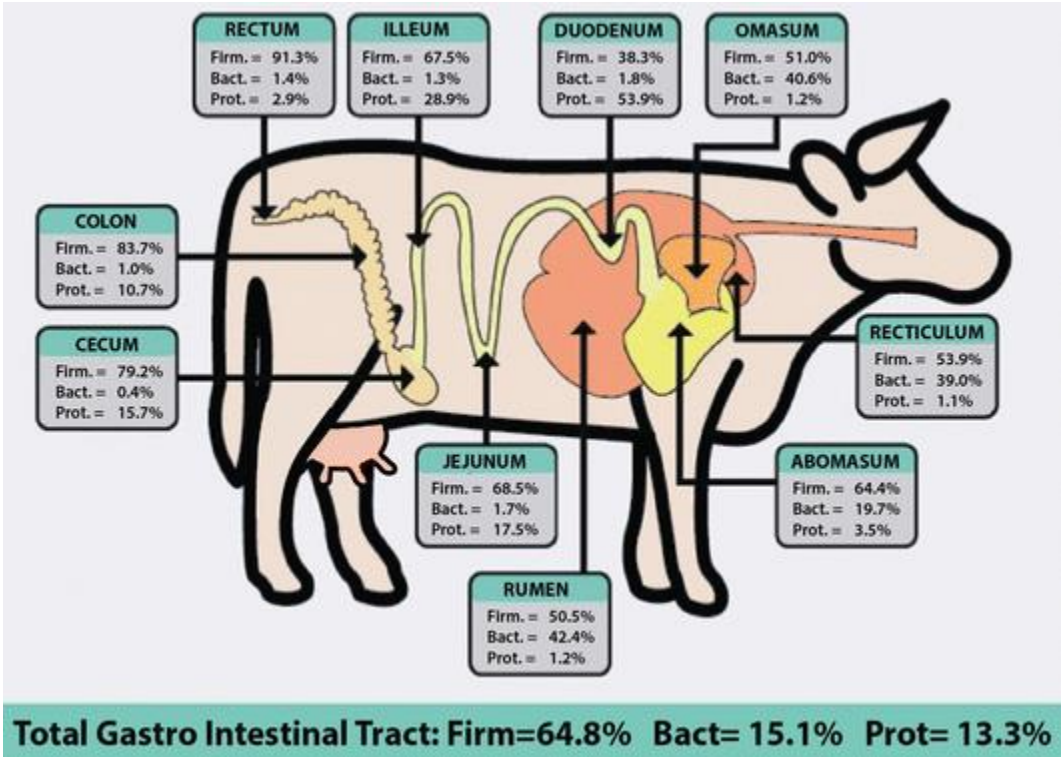

Fig 2.2: Microbial composition of bovine gastro-intestinal tract (GIT). The diagram shows the distribution of the predominant phyla per niche and overall abundances in the GIT, (Dias & Ametaj, 2017).

Apart from the colonizers of the gut most of which are commensals and others mutualistic, bovine faeces is known to harbour a wide variety of microorganisms which can be pathogenic to both animals and humans. These include bacteria such as *Salmonella* spp., *Campylobacter* spp., *Listeria monocytogenes*, *Yersinia enterocolitica* and *Escherichia coli* (Manyi-Loh *et al*., 2016).

2.1.2. Milk microbiota (MM)

Milk harbours a complex microbial community, including microorganisms of industrial importance, which possess health-promoting features and those which are of concern from a food quality or safety perspective (Quigley *et al*., 2013). According to Taponen *et al*., (2019) the mammary gland microbiome and that of the milk can be considered to be highly similar, where the origin of microbes in the milk could be from the upper parts of the mammary gland, but it is very likely that many of these microbes migrate from extra-mammary sites and the environment. From the internal cow teat surface alone, the microbial load has been quantified to be between 10^4 and 10^5 bacterial cells/ml via real

time polymerase chain reaction (qPCR), (Oikonomou *et al*., 2020). The microbial composition of milk is said to be influenced by several different parameters which include among others, in the case of raw milk, the microorganisms present in the teat canal, on the surface of the teat skin, in the surrounding air, in feed as well as other environmental factors including housing conditions, the quality of the water supply and equipment hygiene (Quigley *et al*., 2013; Zhang *et al*., 2015; Derakhshani *et al*., 2018). The number of days in milk postpartum is also said to significantly influence the structure of the microbiome in colostrum and milk (Lindner *et al*., 2011).

2.1.2.1. Composition of colostrum microbiome

Bovine colostrum (BC) is a complex biological fluid that supports the growth and health of the neonates (Lindner *et al*., 2011; Lima *et al*., 2017). It is the first milk produced by the mammary gland in the initial 24 to 96 hours of the postpartum period and its production varies depending on the animal species. In the cow, colostrum secretion continues until the sixth milking (Lindner *et al*., 2011). In addition to the nutrients, colostrum contains valuable microflora, which are widely used as probiotics (Lindner *et al*., 2011; Lima *et al*., 2017).

In a study by Lindner *et al*., (2011) a total number of 29 bacterial strains from bovine colostrum were isolated through traditional culture methods. Culture-dependent 16S rRNA gene sequencing was applied and thirteen cultivable species were identified illustrating the biodiversity present in the colostrum samples. Among them were *Lactobacillus casei* and *Bifidobacterium pseudolongum* with potential probiotic application.

Using culture-independent high throughput NGS of the 16S rRNA gene, Lima *et al*., (2017) characterized the colostrum microbiome in order to determine its potential associations with early-lactation clinical bovine mastitis (BM). The study was a prospective observational type where composite colostrum samples were collected from cows with or without clinical BM during the first 30 days postpartum.

The colostrum core microbiome (defined as the bacterial taxa common to all colostrum samples examined) was composed of 20 taxa. This core taxa included bacterial genera already known to be associated with BM *e.g*., *Staphylococcus*, *Mycoplasma* and *Streptococcus* spp. The samples were dominated by Firmicutes, Bacteroidota, Proteobacteria, Actinobacteriota, Fusobacteria (now Fusobacteriota) and Tenericutes phyla, with the 6 most common taxa listed in descending order of

abundance being *Staphylococcus*; *Prevotella*; unclassified taxa under Ruminococcaceae, Bacteroidales and Clostridiales; and *Pseudomonas* (Lima *et al*., 2017).

2.1.2.2. Composition of milk microbiome

The MM refers to the assemblage of microorganisms present in milk and by extension, microorganism(s) associated with the mammary gland or teat (Oikonomou *et al*., 2020). The mammary gland is infected by various species of bacteria, fungi, algae and more recently identified protozoa and viruses (Bhatt *et al*., 2012; Dhanashekar *et al*., 2012; Motaung *et al*., 2017; Taponen *et al*., 2019; Oikonomou *et al*., 2020). These intramammary infections (IMIs) are the most common cause BM, a condition characterized by the presence of cardinal signs of inflammation *i.e.* pain, swelling and redness with or without heat in infected mammary quarters (Carrillo-Casas and Miranda-Morales, 2012). It is arguably the most important disease for the dairy industry worldwide. It causes significant economic losses due to reduced milk production, discarded milk, premature culling and antibiotic usage (Motaung *et al*., 2017). Approximately 20 to 30% of dairy cows develop clinical mastitis at least once during lactation (Ganda *et al*., 2016). According to Bhatt *et al*., (2012), every dairy herd has cows with subclinical mastitis and the prevalence of infected cows varies from 15 - 75%, whereas the involvement of mammary quarters differs between 5 and 40%. Clinical mastitis is also a serious animal welfare issue as it is associated with pain and reduced well-being of the affected animals (Falentin *et al*., 2016; Motaung *et al*., 2017).

Traditionally udder pathogens have been divided into major and minor pathogens, based on their pathogenicity (Taponen *et al*., 2019). The most common pathogenic genera isolated in mastitic milk samples are *Staphylococcus* (*S.*), *Enterobacteria* and *Streptococcus* (*St.*), which cause the great majority of IMIs (Taponen *et al*., 2019). The primary focus of most subclinical diagnostic programmes is to reduce the prevalence of *St. agalactiae, S. aureus* and other Gram-positive cocci, most notably *St. dysgalactiae* (which may be contagious or be environmentally acquired); and environmental pathogens including *St. uberis, Enterococcus* and numerous other coagulase-negative staphylococci, including *S. hyicus, S. epidermidis, S. xylosus* and *S. intermedius* among others (Bhatt *et al*., 2012).

Other pathogens (which are not by any means least important) that are often implicated in BM cases include *Brucella abortus, Coxiella burnetii, Mycobacterium tuberculosis, M. bovis, M. avium*

subspecies *paratuberculosis, Campylobacter jejuni, Bacillus cereus,* Shiga toxin-producing *E. coli* (E. coli O157:H7), *Listeria monocytogenes, Mycoplasma* spp.*, Salmonella* spp.*, Yersinia enterocolitica* and certain strains of *S. aureus* which are capable of producing highly heat-stable toxins (Dhanashekar *et al*., 2012; FAO, 2014; Abebe *et al*., 2016). Most of the above-mentioned udder pathogens are not only mastitogenic but also have zoonotic potential.

Quite notably from HTS-based studies, there is a common core bovine milk microbiome represented by particular species which are frequently reported across investigations of healthy, subclinical and clinical mastitic cows, ranging in number between 20 - 60 taxa (Bhatt *et al*., 2012; Oikonomou *et al*., 2012 & 2014; Quigley *et al*., 2013; Young *et al*., 2015; Falentin *et al*., 2016; Taponen *et al*., 2019). Derakhshani and colleagues in 2018 compiled a comprehensive review and described this core milk microbiome to be mostly dominated by Firmicutes, Proteobacteria, Bacteroidota and Actinobacteriota as the main bacterial phyla regardless of infection status. As had been previously reported by various authors (Kuehn *et al*., 2013; Oikonomou *et al*., 2014; Addis *et al*., 2016; Falentin *et al*., 2016), members of the following taxa were reportedly common across milk samples: *Staphylococcus*, *Propionibacterium*, *Stenotrophomonas*, *Corynebacterium*, *Pseudomonas*, *Streptococcus*, *Comamonas*, *Bacteroides*, *Enterococcus*, *Lactobacillus*, *Fusobacterium* and unclassified members of Ruminococcaceae and Lachnospiraceae*.* More recently, Oikonomou *et al*., (2020) updated this list to also include *Microbacterium, Aerococcus, Jeotgalicoccus, Psychrobacter* and *Enterobacter.*

This distribution pattern of the MM closely resembles that of the colostrum (Lima *et al*., 2017), the mammary glands and the teat canal and apex (Oikonomou *et al*., 2020). There are however several differences in terms of the relative abundances of the microbes within the different niches as shown on Fig 2.3.

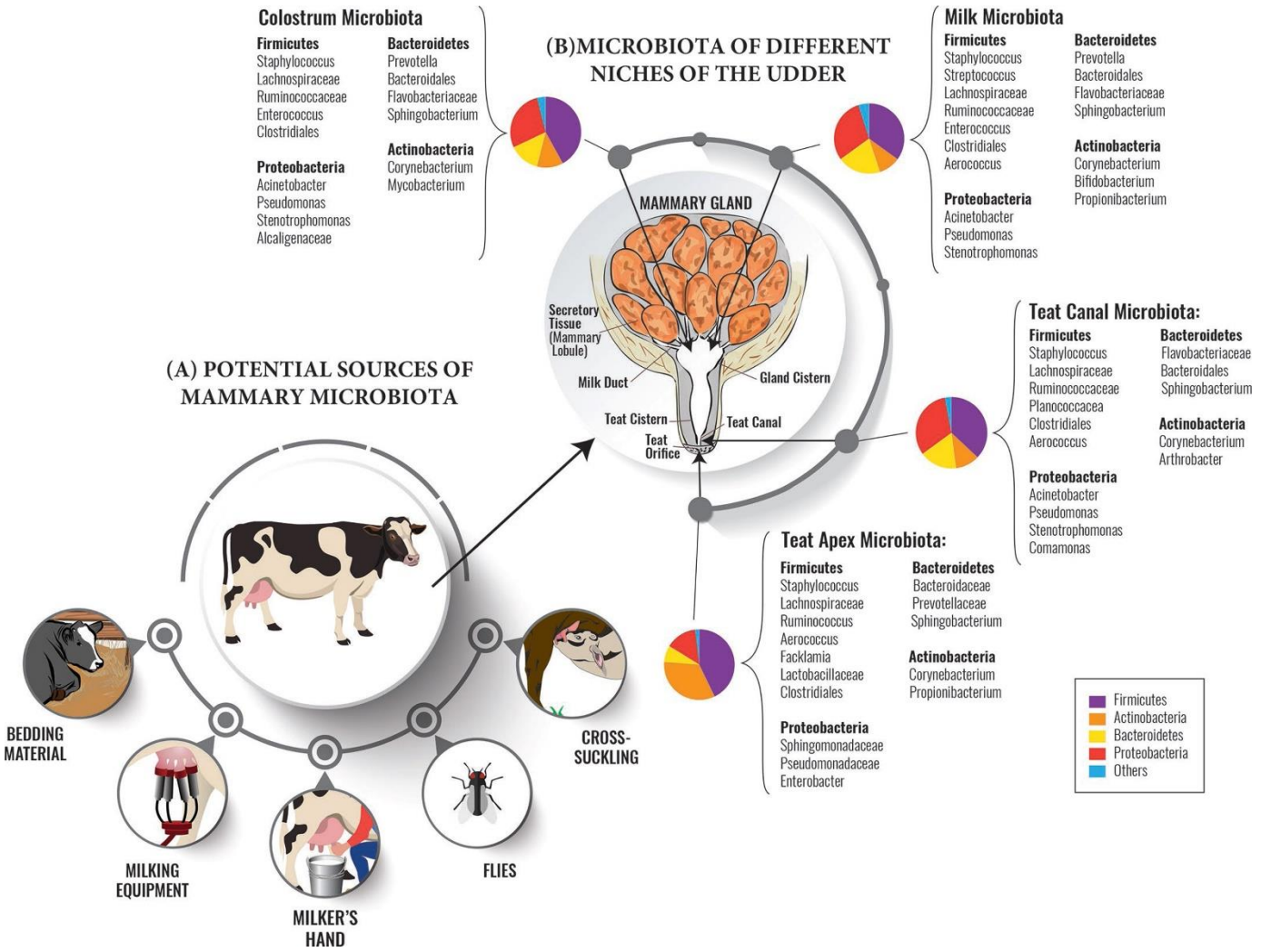

Fig 2.3: Microbial composition of the udder, teats, milk and colostrum. The diagram shows the distribution of predominant phyla in descending order of abundance per niche and the environmental sources of microbes contributing the microbiome structure, (Derakhshani *et al*., 2018).

Of note is that the use of high-throughput DNA sequencing has resulted in several bacterial genera being identified in milk samples for the first time. These include gut associated microbes such as *Bacteroides*, *Faecalibacterium* as well as anaerobic *Prevotella* and *Catenibacterium* (Quigley *et al*., 2013).

2.1.3. Blood microbiota (BM)

There is considerable evidence in the literature that bacteria as well as other microorganisms can reside in blood intracellularly in red blood cells and white blood cells and some circulate freely in the plasma (Potgieter *et al*., 2015; Young *et al*., 2015). While the blood of humans has been thought to be sterile, with the presence of a microbe thought to be indicative of an active infection (Potgieter *et al*., 2015; Velmurugan *et al*., 2020); bovine blood is typically colonized by a myriad of microorganisms which include bacteria, protozoa, viruses, rickettsiae, fungi and helminths (Matjila *et al*., 2008; Berrada and Telford, 2009; Crowder *et al*., 2010). Most of these microorganisms are commensals, many are mutualistic and some are pathogenic (Castillo *et al*., 2019). The pathogenic microorganisms are responsible for a great number of diseases of veterinary significance.

By observation of the PI, most investigations relating to bovine blood are conducted on vector-borne pathogens, transmitted by arthropods such as ticks, lice, mites, fleas and flies. The main focus primarily being on tick-borne pathogens and associated diseases.

There is a wide variety of tick-borne pathogens such as bacteria, rickettsiae, viruses and protozoans that are known to persist in blood (Matjila *et al*., 2008; Sparagano *et al*., 1999). According to Epstein and Price, (2009) these pathogens may cause diseases with varying severity collectively known as tick-borne diseases (TBDs). Examples of major TBDs among livestock in South Africa include anaplasmosis (caused by *Anaplasma centrale* & *A. marginale*), babesiosis (caused by *Babesia bovis* & *B. bigemina*), ehrlichiosis (caused by *Ehrlichia ruminantium*) and theileriosis (caused by *Theileria parva, T. taurotragi, T. mutans, T. annulata* & *T. lestoquardi*) as often reported (Ndlhovu *et al*., 2009; Salih *et al*., 2015).

In general, tick-borne rickettsial diseases (e.g. anaplasmosis and heartwater) and protozoal diseases (e.g. theileriosis and babesiosis) are pre-eminent health and management problems of cattle, small ruminants and buffalo, affecting the livelihoods of farming communities in Africa, Asia and Latin America. These TBDs lead to great economic losses in terms of mortality and morbidity of livestock (Jongejan and Uilenberg, 2004; Sparagano *et al*., 1999; Salih *et al*., 2015). They have a significant impact on meat and milk production and consequently on livestock management (Ndhlovu *et al*., 2009). Some of these pathogens are of zoonotic significance with bacterial pathogens accounting for the largest taxonomic group (42%) of cattle derived zoonotic pathogens (McDaniel *et al*., 2014). Examples of these zoonotic bacterial pathogens include *Coxiella burnetii*, *Rickettsia africae, R. conorii*, *Borrelia burgdorferi*, *Anaplasma phagocytophilum*, *Ehrlichia chafeensis* as well as *Francisella*

tularensis (Kirkan *et al*., 2017). As a matter of fact, recent studies of tick-borne pathogens from different South African provinces have revealed the presence of the above-mentioned zoonotic pathogens (Mtshali *et al*., 2015 & 2017; Halajiani *et al*., 2016; Guo *et al*., 2019; Kolo *et al*., 2020).

Vector-borne pathogens are not only problematic for the livestock industry but are a significant public health concern worldwide. Infections with these pathogens, some of which are emerging, are likely under-recognized due to the lack of widely-available molecular diagnostic techniques. There is therefore an urgent need for further advancement in diagnostic modalities to detect new and known vector-borne blood pathogens (Vijayvargiya *et al*., 2019), and for general exploration of the blood microbiome and pathobiome of both humans and animals.

Other than the tick-borne pathogenic agents, the bacterial agents that have been found in mammalian blood include *Helicobacter pylori*, previously implicated in the development of anemia; *Staphylococcus aureus,* which invades neutrophils and uses them as a means of dissemination of infection; while pathogens such as *Listeria monocytogenes, Salmonella typhimurium* and *Yersinia* spp. are well known to persist intracellularly (Potgieter *et al*., 2015).

As with the gut and milk microbiomes, investigations into the blood microbiome have been propelled by the introduction of NGS techniques to study microbial genetic material present in different body sites. Over the past decade a number of studies have been focused on the establishment of the "healthy" human blood microbiome (HBM) and dispelling the notion that the presence of "foreign" microorganisms in human blood equates to infection or a diseased state, ultimately trying to assign taxonomy and explain the potential origins of the blood residents (Castillo *et al*., 2019). The findings of these studies indicate that the blood microbiome of patients with various conditions and diseases was predominated by members of Proteobacteria, Firmicutes and Cyanobacteria phyla in varying abundances. With the presence of some microorganisms associated with increased risk of occurrence of a particular disease, while presence of others presenting a lower risk of developing the disease (Castillo *et al*., 2019). The NGS-based investigations of the blood microbiome are still at their infancy with the first metagenomic sequencing study recorded in 2016 as can be seen on the timeline on Fig 2.4 (Castillo *et al*., 2019). There is a slow incline in these investigations, with the possibility of the existence of a HBM being met with great disdain among researchers.

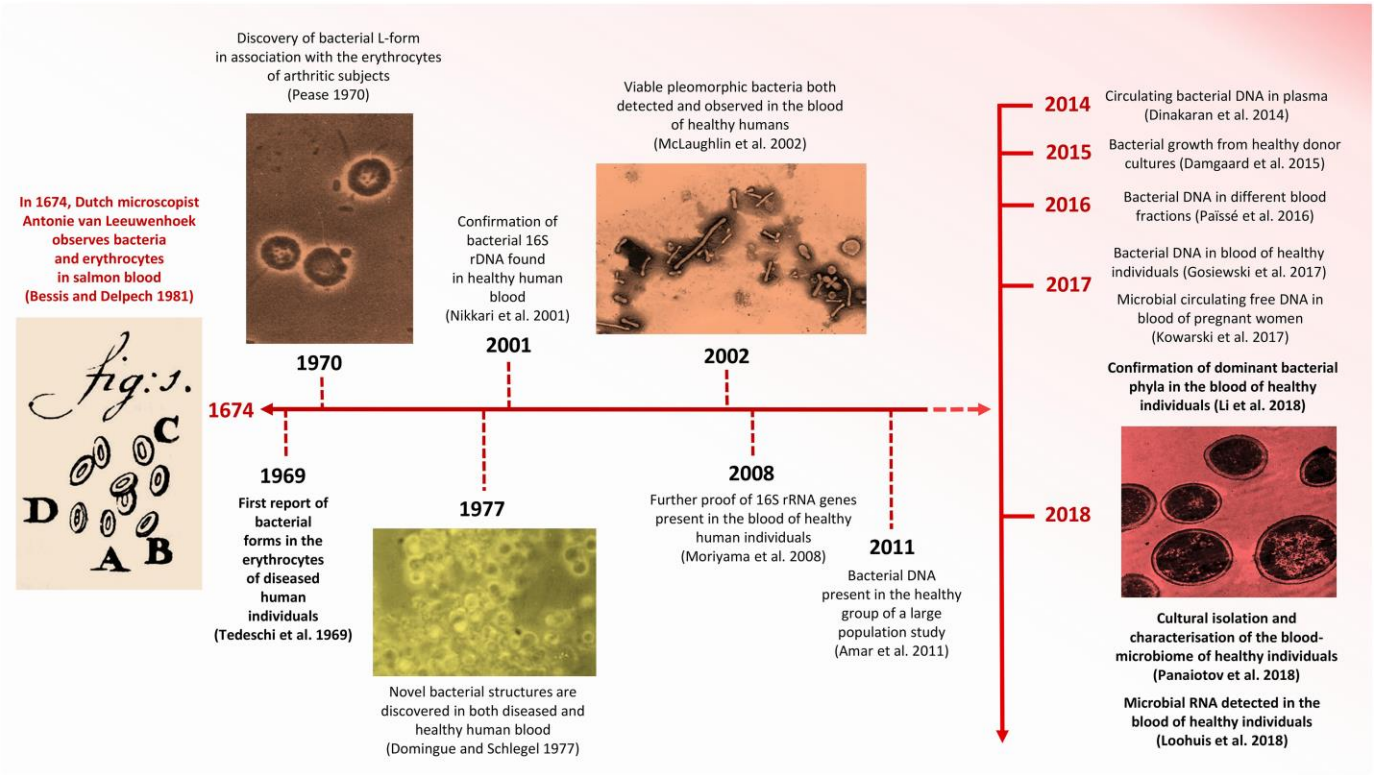

Fig 2.4: Timeline indicating significant advances concerning healthy human blood microbiome (HBM) research, (Castillo *et al*., 2019).

Undoubtedly the human related studies have laid a foundation for the animal based studies (Loohuis *et al*., 2018; Watanabe *et al*., 2018; Whittle *et al*., 2018; Velmurugan *et al*., 2020). Similar to human BM, NGS based studies of animal BMs (although rare) have come about as a consequence of investigating microbial dysbiosis and atopobiosis in populations with a particular disease of interest and never in isolation. A few examples that could be found include profiling of the blood microbiome of broilers to investigate its potential association with Bacterial Chondronecrosis with Osteomyelitis (BCO) in poultry farming (Mandal *et al*., 2016); investigating changes in respiratory microbiota in relation to BM and GM in cats (Vientós-Plotts *et al*., 2017); exploration of the possible migration of GM to blood and milk in cows (Young *et al*., 2015); and investigation of the translocation of GM via the bloodstream to the uterus (Jeon *et al*., 2017).

Results from the above-mentioned investigations showed that bovine white blood cells (macrophages) were mainly comprised of members of Tenericutes (*e.g. Mycoplasma*) which technically now translates to Firmicutes due to reclassification of *Mycoplasma* spp.; Cyanobacteria (*e.g. Streptophyta*); Bacteroidota (*e.g. Prevotella*), Proteobacteria (*e.g. Stenotrophomonas* &

Acinetobacter); Actinobacteriota (*e.g. Micrococcus* & *Kocuria*) and some unclassified bacterial fragments (Young *et al*., 2015). Firmicutes and Cyanobacteria constituted the most predominantly detected bacterial genera in macrophages. Bovine whole blood revealed the presence of Firmicutes (*e.g. Mycoplasma* and *Bacillus*) and Proteobacteria (*e.g. Pseudomonas*) in decreasing order of abundance in Jeon *et al*., (2017)'s study. Furthermore, uterine pathogens such as *Bacteroides*, *Porphyromonas* and *Fusobacterium* were part of the core genera in bovine blood (Jeon *et al*., 2017). Arthropod-borne pathogens have also been characterized in bovine blood including *Anaplasma* spp., *Ehrlichia* spp. and *Bartonella* spp. (Jeon *et al*., 2017, Kolo *et al*., 2020). These studies have formed the basis for subsequent investigations, to which the current study will contribute by adding valuable insights into the putative core microbiome of bovine blood.

2.2. MICROBIAL PROFILING THROUGH SEQUENCING OF THE 16S rRNA GENE

The historical method for performing microbial profiling was dependent on a classical approach, involving culturing the microorganism by preparing a solid or liquid growth medium. Appropriate carbon, energy and electron acceptor sources would be provided depending on the physiological conditions under which the microorganism was to be isolated (Singh *et al*., 2009). Thereafter, it entailed comparison of an accurate morphologic and phenotypic description of type or typical strains with the isolate to be identified, which often proved very difficult (Clarridge, 2004). The cultured cells would also have to be subjected to biochemical testing (Rapp, 2010). These methods have time and again proven to be labour-intensive and time consuming (Rapp, 2010; Hodkinson and Grice, 2014). They have proven problematic for diagnosis of a significant fraction of Bacteria and Archaea (Klindworth *et al*., 2013), with approximately 99% of microorganisms present in certain environments remaining uncultivable by standard culture techniques. Thus, necessitating development of culture independent techniques (Singh *et al*., 2009; Shah *et al*., 2011).

With the technological advancements taking place, a new standard for identifying bacteria began to be developed in the 1980s. It was shown that phylogenetic relationships of bacteria and all life-forms, could be determined by comparing a stable part of the genetic code. Candidates for this genetic area in bacteria include the genes that code for the 5S, the 16S (also called the small subunit), and the 23S rRNA including the spaces between these genes (Clarridge, 2004; Milanese *et al*., 2019). The part of the DNA now most commonly used for taxonomic purposes for bacteria is the 16S rRNA gene (Clarridge, 2004; Chakravorty *et al*., 2007). The 16S rRNA gene is also designated 16S rDNA, and

the terms have been used interchangeably, however the American Society of Microbiology (ASM) policy currently states that “16S rRNA gene” should be used instead (Clarridge, 2004). Sequences from this gene have been used to show the relationship between major branches of life, the Archaea, Bacteria (prokaryotes) and Eukarya (Clarridge, 2004; Milanese *et al*., 2019).

Janda and Abbott, (2007) attribute the usefulness of the 16S rRNA as a housekeeping genetic marker to the following reasons: i) its presence in almost all bacteria, often existing as a multigene family, or operons; ii) the function of the 16S rRNA gene over time has not changed, suggesting that random sequence changes are a more accurate measure of time (evolution); and lastly, iii) the 16S rRNA gene is large enough for informatics purposes. Hodkinson and Grice, (2014) attribute the latter to the presence of a pattern of extremely conserved regions interspersed with hypervariable regions that are widely divergent between different taxa, thus making it easier to identify them with some level of precision.

The 16S rRNA gene sequence is about 1550 base pairs (bp) long and is composed of both variable and conserved regions (Clarridge, 2004; Panek *et al*., 2018). It contains nine hypervariable regions (V1 - V9) that demonstrate considerable and differential sequence diversity among different bacteria (Chakravorty *et al*., 2007; Panek *et al*., 2018). These regions are located at nucleotides 69 - 99, 137 - 242, 433 - 497, 576 - 682, 822 - 879, 986 - 1043, 1117 - 1173, 1243 - 1294 and 1435 - 1465 for V1 to V9, respectively, according to the *E. coli* system of nomenclature (Chakravorty *et al*., 2007). Although no single hypervariable region is able to distinguish among all the bacteria, hypervariable regions V2 (nucleotides 137 - 242), V3 (nucleotides 433 - 497) and V6 (nucleotides 986 - 1043) contain the maximum heterogeneity and provide the maximum discriminating power for analysing bacterial groups (Chakravorty *et al*., 2007; Shah *et al*., 2011). Rausch and colleagues recommend the use of V3 - V4 over other hypervariable regions (Rausch *et al*., 2019).

Of the 20 million sequences deposited on the GenBank database, over 90 000 are of the 16S rRNA gene. This gives a wide variety of previously deposited sequences with which to compare the sequence of an unknown strain. In general, the comparison of the 16S rRNA gene sequences allows differentiation between organisms at the genus level (>90%) across all major phyla of bacteria, in addition to classifying strains at multiple levels, including species and subspecies level (65 - 83%), with about 1 - 14% of the isolates remaining unidentified after testing (Clarridge, 2004).

In the recent past, the most commonly used approach to microbial profiling and species identification was cloning of vectors such as cosmids, fosmids or Bacterial Artificial Chromosomes (BACs) and

sequencing of the 16S rRNA gene using conserved broad-range PCR primers (Chen & Pachter, 2005; Singh *et al*., 2019; Klindworth *et al*., 2013).

Direct sequencing of PCR amplicons became feasible with the establishment of new massively parallel technologies (Klindworth *et al*., 2013; Kwong *et al*., 2015). These early methods, although technically challenging as they relied on direct sequencing of RNA or sequencing of reverse transcription-generated DNA copies, have aided in understanding the genetic diversity, population structure and ecological roles of the majority of microbial organisms (Singh *et al*., 2009).

2.2.1. Next-generation sequencing (NGS)

As a general consensus from observation of the PI, metagenomics (the principle), has frequently been coupled with HTS technologies when expounding on its application *e.g.* shotgun metagenomics sequencing; metagenomics pyrosequencing; 16S rRNA metagenomics sequencing; *etc*. Thus, it is imperative to elucidate the role and influence of metagenomics in development of high throughput NGS technologies prior to delving into their application in the field of microbiology for exploratory microbial community profiling research.

2.2.2. Metagenomics in microbial profiling

First applied in environmental studies around 1998, the broad field of metagenomics is mainly referred to as environmental genomics, ecogenomics or community genomics (Wooley & Ye, 2010). The field is said to stem from the culture-independent retrieval of 16S rRNA genes (Chen & Pachter, 2005). By definition, metagenomics is a direct genetic analysis of genomes contained within an environmental sample (Shah *et al*., 2011; Thomas *et al*., 2012). Unlike classical microbial community profiling which relies on cultivated clonal cultures, this method does not require laboratory cultivation and isolation of individual species (Wooley & Ye, 2010; Shah *et al*., 2011).

To elaborate, metagenomics works by direct extraction and cloning of DNA and subsequent genomic analysis of microorganisms from their natural environment for example, sea water and soil (Singh *et al*., 2009). It applies modern genomic techniques and bioinformatics tools to directly access the genetic content of entire communities of organisms (Thomas *et al*., 2012). The method typically

encompasses two particular sequencing strategies: i) amplicon sequencing, most often of the 16S rRNA gene as a phylogenetic marker or; ii) shotgun sequencing, which captures the complete breadth of DNA within a sample (Salipante *et al*., 2014; Watanabe *et al*., 2018; Rausch *et al*., 2019).

As opposed to shotgun sequencing of genomic DNA extracted from a sample wherein random fragments of bacterial genomes (including host DNA or other microorganisms present) are sequenced and classified, 16S rRNA amplicon sequencing can be targeted specifically against the organisms of interest *e.g.* bacteria, does not require the availability of reference genome sequences, and can be employed in cases where only trace amounts or poor-quality bacterial DNA templates are available (Salipante *et al*., 2014).

Metagenomics soon gained popularity after its application revealed that there was a vast majority of microbial biodiversity that had been missed by cultivation-based methods (Wooley & Ye, 2010), since about 99% of all microorganisms remain uncultivable in growth media to date (Singh *et al*., 2009).

It soon became a method of choice being applied in a wide range of metagenomic profiling studies. Due to its realised ability to uncover an enormous functional gene diversity in the microbial world around us in later studies, the field of metagenomics has been responsible for substantial advances in microbial ecology, evolution, and diversity over the past 5 to 15 years. It has further influenced the explosive development of state-of-the-art, HTS technologies and the advancement of bioinformatics tools (Wooley & Ye, 2010; Shah, 2010; Thomas *et al*., 2012; Grada & Weinbrecht, 2013; Hodkinson & Grice, 2014; Kwong *et al*., 2015). The principle has further lead to the discovery of novel biocatalysts or enzymes, genomic linkages between function and phylogeny for uncultured organisms, and evolutionary profiles of community function and structure. It can also be complemented with metatranscriptomic or metaproteomic approaches to describe expressed activities (Singh *et al*., 2009; Shah *et al*., 2011; Thomas *et al*., 2012). Metagenomics is also a powerful tool for generating novel hypotheses of microbial function (Thomas *et al*., 2012) and may be useful for pathogen discovery and identification (Oikonomou *et al*., 2012).

Based on these attributes, metagenomics has been adopted in many research laboratories where it is used to study the microbial composition of clinical samples. Because the method provides access to the functional gene composition of microbial communities, it thus gives a much broader description than phylogenetic surveys which are often based only on the diversity of one gene (Shah *et al*., 2011; Thomas *et al*., 2012). It is therefore envisaged that in the future, the method will be used in the same manner as 16S rRNA gene fingerprinting methods to describe microbial community profiles and will

therefore become a standard tool for many laboratories and scientists working in the field of microbial ecology (Thomas *et al*., 2012).

2.2.3. Next-generation sequencing: the principle, technologies and application

2.2.3.1. Principle of NGS and associated technologies

Due to advancements in metagenomics, the initial traditional (first-generation) sequencing techniques and platforms evolved into second-, third- and fourth-generation, more commonly referred to as next-generation sequencing (NGS) (Grada and Weinbrecht, 2013; Reuter *et al*., 2015).

NGS platforms perform massively parallel sequencing, during which millions to billions of fragments of DNA from a single sample are sequenced in unison (Grada and Weinbrecht, 2013). Massively parallel sequencing technology facilitates HTS, where vast quantities of data requiring modern computation methods are utilised to assemble the sequence reads. These reads allow an entire genome to be sequenced in less than one day (Grada and Weinbrecht, 2013; Kwong *et al*., 2015).

The different NGS platforms are all unique and are based on different technologies but typically follow a general pattern or sequence of steps. General steps involved in the DNA sequencing using NGS are: i) library preparation (by random fragmentation of genome and ligation with appropriate adaptors); ii) amplification of library; and iii) sequencing using different approaches (*i.e.* pyrosequencing, sequencing by ligation or sequencing by synthesis). The steps are shown on Fig 2.5 (Gupta and Verma, 2019). The different sequencing pathways may include one or more of the following: i) template preparation, emulsion PCR and sequencing data analysis; ii) template preparation, sequencing and imaging and data analysis; iii) template amplification, sequencing and imaging and data analysis; iv) template preparation, sequencing and imaging and alignment to a reference genome or; v) DNA fragmentation, sequencing and data analysis (Grada and Weinbrecht, 2013; Hodkinson and Grice, 2014). The specific strategy employed by each platform determines the quality, quantity and biases of the resulting sequence data and the platform's usefulness for particular applications (Reuter *et al*., 2015). However, the particular questions being addressed in a particular investigation will guide the experimental design and the methodology for generating, processing, and interpreting data (Hodkinson and Grice, 2014).

The different NGS technologies are classified based upon the type of technology, chemistry (sequencing by ligation, by synthesis or nanopore), detection system (optical, solid state or electrical detection), and method of amplification (by emulsion PCR, bridge PCR or no amplification needed) used in different generations of sequencing platforms (Gupta & Verma, 2019).

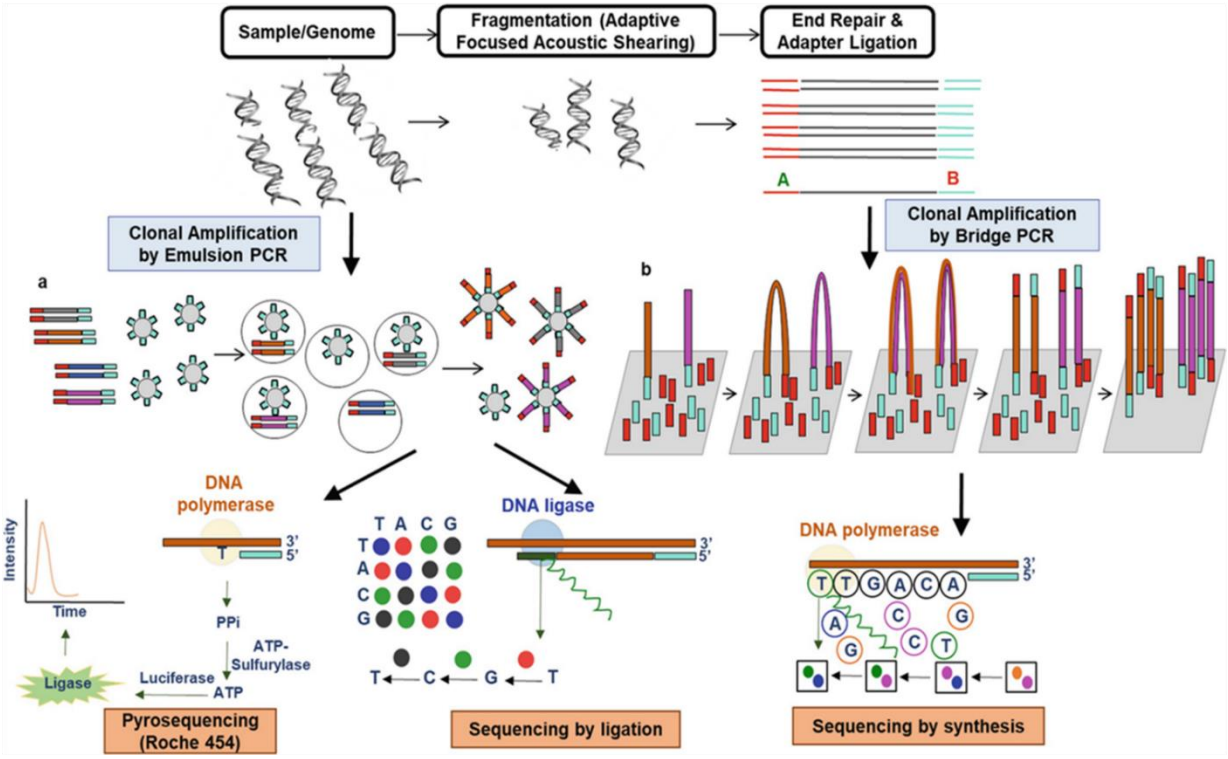

Fig 2.5: Schematic representation of the basic steps involved in DNA sequencing using different NGS platforms (Gupta & Verma, 2019).

The chain termination sequencing method published by Sanger and colleagues in 1977, along with Maxam-Gilbert method of DNA sequencing are considered today as the traditional or first -generation sequencing methods (Sanger *et al*., 1977; Gupta & Verma, 2019).

The challenge most often encountered with Sanger sequencing was its limitation in read length (<1000 bases) and the associated high cost per base (Gupta & Verma, 2019). The search for more efficient methods for sequencing long, complex pieces of DNA such as entire chromosomes at lower cost, prompted the need for development of HTS technologies (Grada & Weinbrecht, 2013; Kwong *et al*., 2015; Reuter *et al*., 2015). The timeline for the sequencing events and introduction of platforms of different generations of sequencing technologies is shown Fig 2.6.

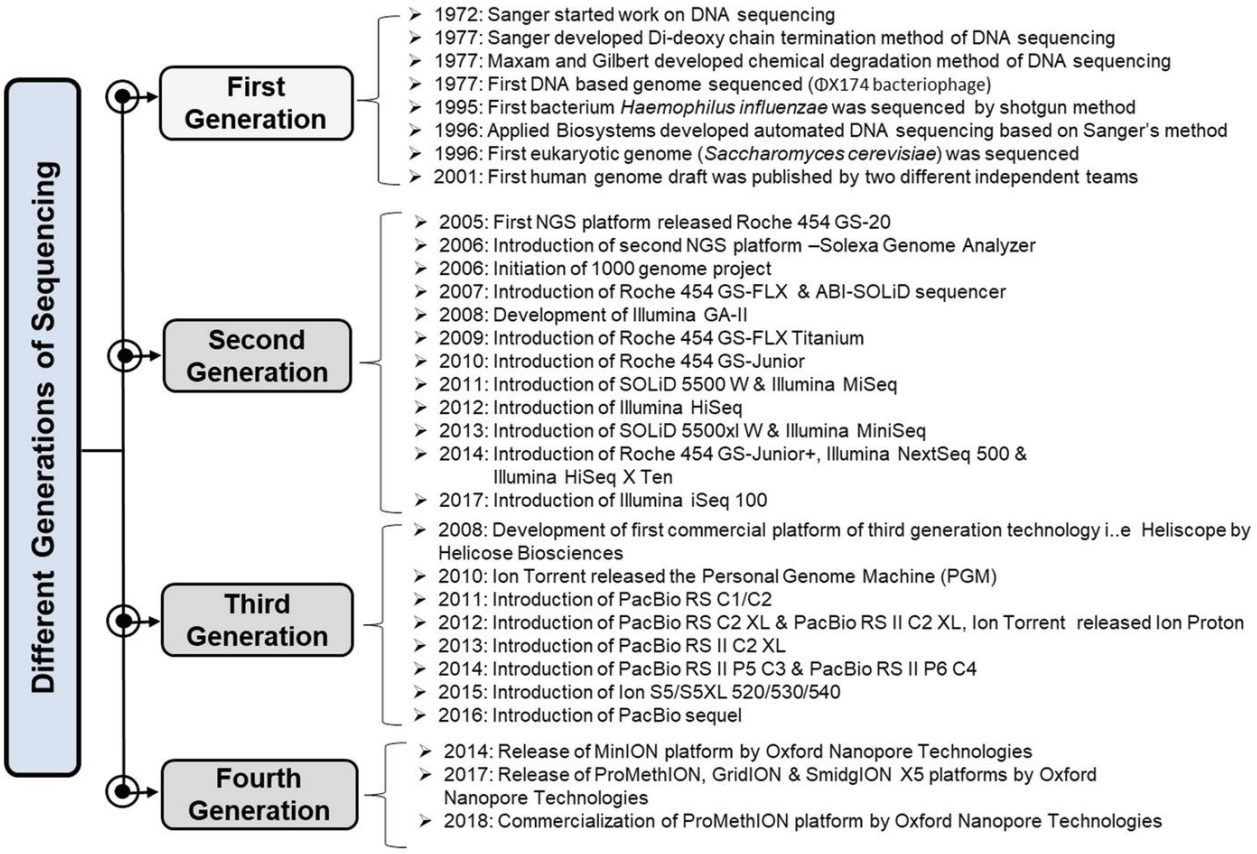

Fig 2.6. Classification of the next generation sequencing (NGS) platforms and timeline of sequencing events, developments, and introduction of different generations of sequencing platforms, (Gupta & Verma, 2019).

Initial technological advances focused on enhancing Sanger sequencing. These modifications included fluorescent labelling of molecules, development and utilisation of capillary-based instruments, and automation of these processes to allow analysis of multiple samples in parallel (Kwong *et al*., 2015).

Initially described in 1979, shotgun sequencing, where longer segments of DNA were randomly fragmented into smaller segments for Sanger sequencing, was an early step towards facilitating whole genome sequencing, but was slow and labour-intensive for an entire genome, requiring a map to assemble the sequenced fragments. Shotgun sequencing was later improved and it has been incorporated into newer high throughput NGS methods for a variety of applications including whole genome sequencing (WGS; Kwong *et al*., 2015).

Only in 2006 did the first HTS technology, Roche's 454 GS 20, become successfully applied for large scale biodiversity analysis and was key to uncovering the 'rare biosphere' (Grada & Weinbrecht, 2013). One shortcoming of the 454 approach is that it frequently misidentifies the length of homopolymers, which are stretches of nucleotides in which all bases are identical. Additionally, this technology is often considered to be cost ineffective to newer platforms (Hodkinson & Grice, 2014). Although the technology broke new ground when it was introduced, 454 Life Sciences no longer supports the platform and they have reportedly closed sequencing operations and ceased production (Hodkinson & Grice, 2014; Kwong *et al*., 2015; Gupta & Verma, 2019). Currently, the most advanced variety of the sequencing platforms within this family *i.e.* GS FLX+ System with the GS FLX Titanium Sequencing Kit XL+ can produce 1 million reads per run with reads up to 1000 bases in length (Hodkinson & Grice, 2014). The platforms can also facilitate *de novo* assembly (Kwong *et al*., 2015).

With the ever-increasing trends in development of new technologies, today read lengths of between 150 - 300 bp can be obtained from such platforms, termed second-generation sequencing platforms (Gupta & Verma, 2019). Over time the platforms are becoming available for use in diagnostics and due to the reduced cost per base price and greater sequencing depth, microbiologists are able to access the tools to use in improvement of disease diagnosis and understanding the causative agents (Grada & Weinbrecht, 2013; Klindworth *et al*., 2013).

One of the latest inventions introduced in 2006, the Illumina platform has attracted most researchers due to its cost effectiveness and comparatively high sequencing depth, despite having a limitation of short read lengths (between 400 - 700 bp) than traditional Sanger sequencing with read lengths of approximately 750 bp (Hodkinson & Grice, 2014; Reuter *et al*., 2015; Gupta & Verma, 2019). However, this limitation is compensated for by the much larger number of sequence reads generated. It is reported that in 2009, pyrosequenced metagenomes could only generate between 200 - 500 megabases, while Illumina platforms could generate around 20 - 50 gigabases, however throughout the years these outputs have increased by a great order of magnitude (Klindworth *et al*., 2013; Grada & Weinbrecht, 2013). Currently, the longest reads produced on an Illumina platform can be found on the MiSeq, which can produce paired-end reads that are 300 bases in length each (Hodkinson & Grice, 2014), this making it an ideal choice of technology for the current study. Another recent release is the NextSeq 500, which is being marketed as the first high-throughput desktop sequencer (Hodkinson & Grice, 2014). The platform with the greatest output overall is the HiSeq 2500, producing 4 billion fragments in a paired-end fashion with 125 bases for each read in a single run (Hodkinson & Grice, 2014), although it is said to be better suited to human genome sequencing (Kwong *et al*., 2015). Illumina has recently released the HiSeq X Ten, which is an array of 10 HiSeq machines sold as a unit, for higher throughput than ever before (Hodkinson & Grice, 2014; Reuter *et al*., 2015). The

Illumina platforms generally produce high throughput data at low cost per output, they are suitable for microbial genomes although may still be limited by short read lengths (Kwong *et al*., 2015).

Similarly, Ion Torrent Personal Genome Machine (PGM) has relatively low costs and rapid sequencing speed. On the other hand, Pacific Bioscience (PacBio) now employs the single-molecule real-time (SMRT) and Single-Molecule Long-Read (SMLR) sequencing technology, designed to achieve average read lengths of more than 3000 bp and 10-20 kb, respectively (Kwong *et al*., 2015; Gupta & Verma, 2019; Jeong *et al*., 2021). These technologies, referred to as third generation sequencing platforms have allowed researchers to reach milestones and opened a new dimension in biodiversity analysis (Kwong *et al*., 2015; Gupta & Verma, 2019).

Fourth-generation sequencing platforms have also been introduced to the market between 2014 and 2018. Although still at their infantile stages, they have been hailed for their very low cost, ability to generate read lengths of >150 000 bp and to directly sequence the fixed tissue and cells by using second-generation methodology coupled with nanopore technology (Gupta & Verma, 2019). The method enhances the rate of fragment capture a thousandfold, thus with this approach, determination of expression of large number of genes in the cell is possible in parallel fashion for several types of RNA. Furthermore, nanopore technologies generate large amounts of data at a very short period, however their major drawback of is their higher error rates compared to all other technologies mentioned (Gupta & Verma, 2019). There are without a doubt, many more inventions and improvements to the current cohort of high throughput NGS technologies on the pipeline.

The major advantage of these NGS platforms is that they do not require cloning of the DNA before sequencing, thus eliminating one of the main biases associated with clinical and environmental sampling (Singh *et al*., 2009; Shah *et al*., 2011; Panek *et al*., 2018). However, the diversity of technologies creates added complexity to the experimental design of studies and most importantly the data analysis and interpretation (Klindworth *et al*., 2013).

With the continual modification of sequencing technologies and improvement of bioinformatics tools, the limitations of using culture-independent molecular techniques are becoming negligible. These tools have the potential to offer some insight in the microbial communities present in various types of body fluids and sites including faeces, milk, and blood and increase our understanding of their taxonomy and functionality. Sequencing and analysis of hypervariable regions within the 16S rRNA gene can provide relatively rapid and cost-effective methods for assessing bacterial diversity and abundance and may be useful for pathogen identification and discovery (Oikonomou *et al*., 2012; Thomas *et al*., 2012).

Once sequences have been retrieved from the sequencer a variety of softwares implementing different pipelines can be used to turn paired-end FASTQ format files into merged, denoised, chimera-free, inferred sample sequences (Callahan *et al*., 2016). The bioinformatics tools are easily available on the internet with step by step guides on how to perform the series of commands into a terminal window once the open-source softwares have been downloaded and ran onto a hardware device (Kuczynski *et al.*, 2011; McMurdie & Holmes, 2013).

The softwares are laden with tools for statistical analysis of phylogenetic sequencing data within the programming environment, making it easy to read the data output of several of the most common operational sequence unit (OTU) or amplicon sequence variant (ASV) clustering pipelines. The use of ASVs has been reportedly found to allow for greater precision and reproducibility in taxonomic assignment compared to the use of OTUs when using the same sequencing data (Callahan *et al*., 2016; Kolbe *et al*., 2019). Comparison of ASVs and OTUs in 16S rRNA sequence data analysis has previously shown that despite the larger number of OTUs generated as opposed to the number of ASVs from the same sequence data, similar trends could be seen in plots of observed OTUs/ASVs for alpha diversity analysis (Capitunan, 2018). However, the ability to distinguish sequence variants differing by as little as one nucleotide, imperceptible to OTU methods, makes the ASV inferring method preferable (Callahan *et al*., 2015; Capitunan, 2018; Farrell *et al.*, 2019). The precision of this method improves downstream measures of diversity and dissimilarity and potentially allows amplicon methods to probe strain-level variation (Callahan *et al*., 2019). The softwares also represent this data in a unified, integrated form amenable to many modern analysis methods (McMurdie & Holmes, 2013). The microbial community data can then be viewed and analysed as graphical and textual output (Kuczynski *et al.*, 2011). With these integrated means of data representation it becomes easy to use methods such as canonical correspondence analysis, discriminant correspondence analysis, sparse linear discriminant analysis, etc. to explain clinical or environmental response variables (McMurdie & Holmes, 2013). Furthermore, these tools employ a scalable matrix-based visualization to show intersections of sets, their size, and other properties (Conway *et al*., 2017).

2.2.3.2. Application of NGS in microbial profiling

The rapid and substantial cost reduction in NGS has dramatically accelerated the development of sequence-based metagenomics, leading to a massive increase in the metagenomics sequence datasets in the past few years across many fields of research (Wooley & Ye, 2009; Shah *et al*., 2011;

Thomas *et al*., 2012; Grada & Weinbrecht, 2013; Hodkinson & Grice, 2014; Kwong *et al*., 2015). Due to the ability of HTS to produce large 16S rRNA datasets that contain hundreds of thousands of 16S rRNA fragments, it has enabled deep views into hundreds of microbial communities simultaneously and revealed much greater species diversity in many environments *e.g*., soil, ocean water and human bodies, than previously anticipated (Shah *et al*., 2011; Oikonomou *et al*., 2012 & 2020).

Metagenomics projects have very broad applications, from ecology and environmental sciences, to the chemical industry and human health (*e.g*. the human gut microbiome metagenomics). More recently, metagenomics via HTS has been applied in animal health studies to elucidate the microbiome of the gut and milk, and in a few cases the microbiome of blood. Other than the large-scale shotgun metagenomics, there are small-scale approaches such as the 16S rRNA-based surveys, and targeted metagenomics (to study the microbes in their environments) which are more commonly used (Wooley & Ye, 2009).

For the purpose of this study NGS technologies (and associated metagenomics) were investigated for their utility and application in establishment of the gut, milk and blood microbiomes of bovines and other mammals, including humans (where these technologies have seen greater application).

From the literature investigation it was found that various NGS platforms have been utilised to characterize the diversity of bacterial communities in human and animal gut. NGS technologies utilized include sequencers from Roche, Thermo Fischer Scientific, Pacific Bioscience and Illumina to describe the microbiome of the gut (Panek *et al*., 2018). Majority of these HTS studies conducted to date are based on the 454 Roche pyrosequencing and the Illumina platforms (Zeineldin *et al*., 2018). The studies include profiling of the gut microbiome of infants (Ruiz *et al*., 2019; De Leoz *et al*., 2015) and adult humans (Rampelli *et al*., 2013; Panek *et al*., 2018). Chiefly, the bovine studies include investigations into the microbial community of the rumen due to its importance in FE and contribution to milk and meat production. Such studies include investigation of the potential role of the bovine rumen microbiome in modulating milk composition and FE by Jami *et al*., (2014) and Zhong *et al*., (2018) as well as translocation of GM to blood and milk and its role in neonatal immune imprinting (Young *et al*., 2015). Franco-Lopez *et al*., (2020) attempted to define the correlation of bacterial genera in the rumen with vitamin B12 abundance in milk. Malmuthuge and coworkers provide a review of the gut microbiome and its potential role in the development and function of newborn calf gut. They elaborate on the pre-ruminant and ruminant gut colonization during and postpartum and highlight the role of HTS platforms in the description of the gut microbiome (Malmuthuge *et al*., 2015). Similarly, McSweeney and Mackie, (2012) provide a lengthy review of the rumen microbiome and

delve into the use of omics approaches to understanding rumen microbial function. There have been attempts of studying the hindgut, where an investigation of the composition of the perinatal intestinal microbiota in cattle was conducted (Alipour *et al*., 2018). Microbial analysis of bovine faeces using NGS technologies have been very limited (Dowd *et al*., 2008; Oikonomou *et al*., 2014; Young *et al*., 2015; Franco-Lopez *et al*., 2020, Koester *et al*., 2020).

The technologies have also been applied in milk and colostrum investigations (Perez *et al*., 2007; Ruiz *et al*., 2019; Oikonomou *et al*., 2020) where they were found to have been efficient in uncovering a much greater diversity of bacteria than what has been previously reported in both culture-based and other culture-independent investigations. The potential effects of the milk microbiome on infant health and imprinting of the neonatal gut have also been investigated using these technologies (Perez *et al.,* 2007; Ruiz *et al*., 2019).

A number of researchers in the animal science related fields adopted its use in investigations of bovine milk (Kuehn *et al*., 2013; Oikonomou *et al*., 2012 &2014; Quigley *et al*., 2013; Young *et al*., 2015; Ganda *et al*., 2016; Franco-Lopez *et al*., 2020) and colostrum (Lima *et al*., 2016) microbiomes. The bovine milk studies were mostly based on discovering the microbiome of mastic milk in both clinical and subclinical cases (Oikonomou *et al*., 2012 & 2014; Ganda *et al*., 2016); in another case NGS technology was used to describe the microbiome of milk from healthy cows (Derakhshani *et al*., 2018); while a few researchers also used NGS for screening of both healthy and mastitic cow milk (Kuehn *et al*., 2013; Falentin *et al*., 2016; Hoque *et al*., 2019; Taponen *et al*., 2019). Metagenomic pyrosequencing via 454 Roche platform was applied in another study to analyse the milk microbiome of two species of cattle and their cross-breeds to investigate if breed type has an effect on the milk microbial composition in subclinical mastitis cases (Bhatt *et al*., 2012).

One study applied Illumina sequencing of the 16S rRNA to assess the effect of disease and antibiotic therapy on the milk microbiome (Ganda *et al*., 2016). Another group of researchers conducted a theory driven investigation to prove a hypothesis that pathogens follow an entero-mammary pathway from the gut where they are trafficked by macrophages to the mammary glands via the bloodstream (Young *et al*., 2015). Furthermore, HTS of the 16S rRNA gene has been applied in a study by Quigley *et al*., (2013) to assess the microbial population of milk from a selection of commercial milk producers, pre- and post-pasteurization with the aim of addressing the potential bias of using culture dependent techniques for screening of milk. The technology has also been applied in various studies to generate milk microbiome profiles of ruminants including goats, sheep, donkeys, buffaloes, water deer and reindeer (Oikonomou *et al*., 2020).

There is a definite paucity of studies based on microbial profiling of blood. There are a handful of reports on screening of human blood using metagenomics and amplicon targeted sequencing via various NGS technologies. They mostly used Illumina Miseq platforms according to the review by Castillo *et al*., (2019). The main aim of some of these studies was to evaluate the methods' efficiency as a diagnostic tool in patients with various blood infections rather than profiling of the microbiome contained within blood. It was used to screen samples of patients with sepsis in comparison to blood samples spiked with pathogens of interest (Faria *et al*., 2015; Watanabe *et al*., 2018); as well as a confirmatory tool in samples whose infection status with vector-borne bacteria, protozoa & helminths had been established using other methods (Vijayvargiya *et al*., 2019); furthermore NGS has been applied in evaluation of the biological effects associated with blood and bone marrow transplantation (Chapman *et al*., 2012). In relation to bovine blood a study based on white blood cells was conducted by Young *et al*., (2015), in conjunction with microbial profiling of milk and faeces. Another study was conducted to investigate the role of blood in translocation of uterine microbes from the gut to the uterus (Jeon *et al*., 2017).

Collectively, the sequencing platforms utilized in the above-mentioned studies were Illumina HiSeq 2500 & MiSeq, Roche 454 and less frequently Ion Torrent PGM and Pac Bio. Historically, many 16S rRNA amplicon sequencing experiments were performed by using Roche 454 massively parallel pyrosequencing, both because it was the first commercially available system and because it later offered the longest read lengths, permitting interrogation of a larger and consequently more informative fraction of the 16S rRNA gene according to Salipante *et al*., (2014). However, this platform is currently being phased out by the manufacturer (Hodkinson & Grice, 2014; Kwong *et al*., 2015). At present, Illumina and Ion Torrent PGM are the highest selling platforms. Of note is that Illumina Miseq comes across as a preferred platform for sequencing, yielding the desired results as opposed to Ion Torrent PGM. Comparative studies of microbial community profiles where both platforms' efficacy was evaluated showed that Ion Torrent PGM platform had higher error rates and prematurely truncated sequence reads, resulting in organism-specific biases compared to Illumina platforms (Salipante *et al*., 2014; Panek *et al*., 2018). This therefore makes Illumina platforms more desirable in sequencing and for application in microbial community profiling. Thus, in addition to selection of appropriate sampling techniques and targeting the appropriate hypervariable region to sequence, another component to consider in microbial profiling research is the choice of sequencing platform as it has been elucidated that the type of NGS platform used has the potential for differential bias in bacterial community profiling.

CHAPTER 3: MATERIALS AND METHODS

3.1. RESEARCH DESIGN

This study followed an exploratory experimentation concept. It employed a mixed-method complex approach, incorporating both qualitative & quantitative methods of data collection & interpretation. Fig 3.1 shows the schematic representation of the research design from sample collection to analysis.

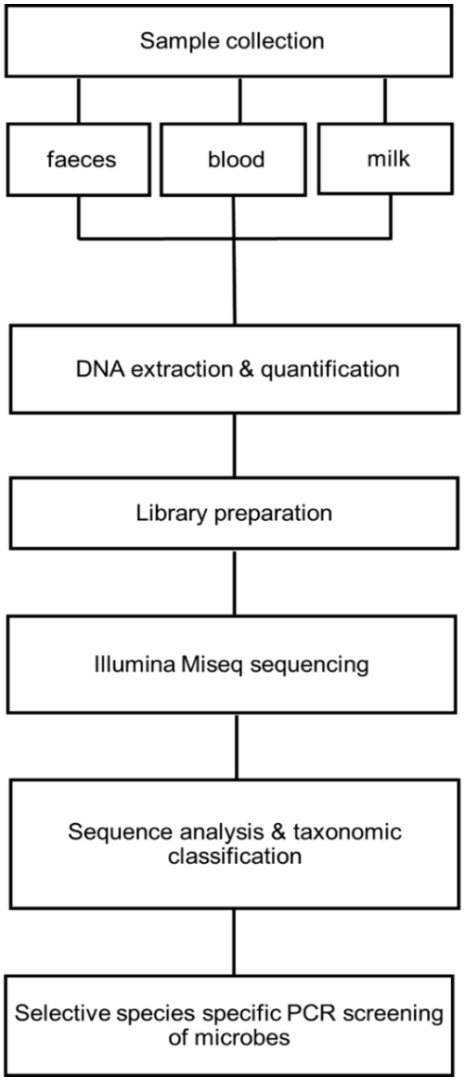

Fig 3.1: Schematic representation of the research design from sample collection to sample analysis.

The study was exempted (NWU-01757-20-A9) by the North West University's Faculty of Natural and Agricultural Sciences Research Ethics Committee and sampling was approved by the Animal Health

Unit of the Department of Agriculture Forestry and Fisheries under Section 20 of the Animal Disease Act of 1984 (Act 35 of 1984), Ref no:12/11/1/3 (887). The cattle owners and herders verbally agreed to an informed consent to participate in the study.

Sampling was conducted in Waaihoek, a rural area situated on the outskirt of Ladysmith in the north-western boundary of KwaZulu-Natal province, in consultation with uThukela Veterinary Services under the Ladysmith State Veterinary Animal Health Division. The animals were gathered at a local cattle dip site (GPS coordinates: -28.46822280; 30.0880990) which serves the communal farmers of Niekerskraal and Waaihoek, under the uThukela District Municipality. Sampling occured on the 17th of April 2019 at 8:00 am. The map (Fig 3.2) shows outlines of South Africa and KwaZulu-Natal province indicating the major towns and metropolitans as well as the GPS coordinates of the dip site located within Patronella Farm 14026 (green pentagonal shape). For the purpose of this study the sampling site will hereinafter be referred to as Waaihoek.

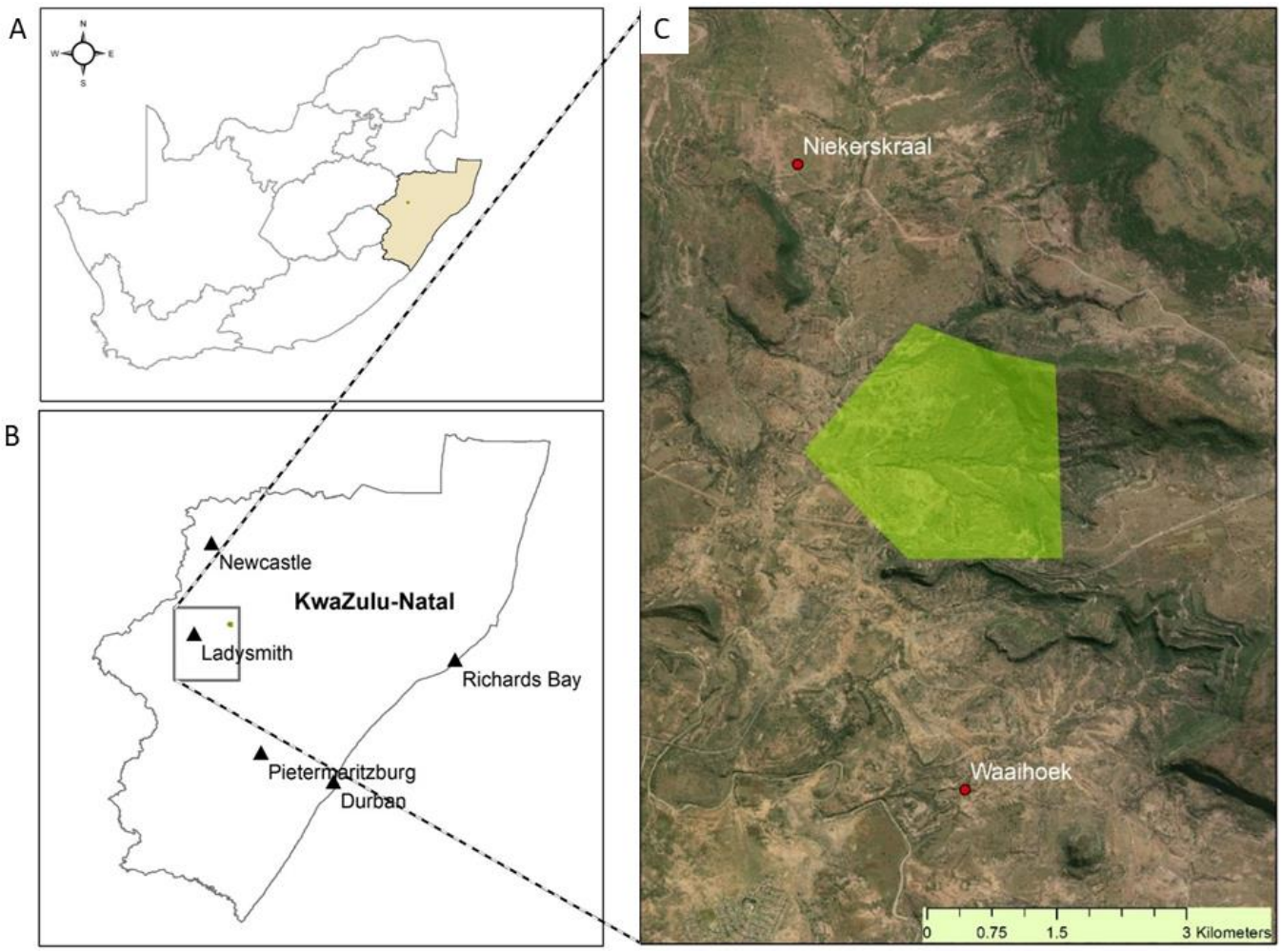

Fig 3.2: Map of South Africa (A) and KwaZulu-Natal (B) showing the location of the dip site, situated between Niekerskraal and Waaihoek in the uThukela district, Created using ESRI ARC GIS Desktop, ArcMap v 10.4. The sampling site is shown with pentagon over the Google Earth image (C).

3.2. RESEARCH METHODOLOGY

3.2.1. Sample collection and processing

Sampling of faeces, milk and blood was achieved through the aid of certified Animal Health Technicians as well as the farmers and / or herders who assisted in restraining of the cattle into the crush pens. The cows were not placed on any special diet prior to sampling however, the owners and / or herders confirmed that their daily diet is (was) typically constituted of kitchen left overs in the morning and during the day they are (were) allowed to roam in search of forage, either supervised or unsupervised. At night they are (were) housed in kraals situated in close proximity to the homes of the owners. Prior to sampling the cows were rested and allowed to forage on the overgrown thatch grass surrounding the dip tank and crush pens. The sampling population included n = 110 mixed breed cows, *i.e.* 33 lactating and 77 non-lactating (dry) which were representative of the entire cow population owned by the Waaihoek community utilising the Niekerskraal cattle dip. Three sample sets (*i.e.* milk, blood and faeces) were collected per lactating cow while only blood samples were collected from the non-lactating cows.

While restrained, a handful (~ 3 grams) of faecal sample was collected directly from the rectum of each of the cows with gloved hands as described by Gibbons *et al*., (2020). Briefly, a gloved hand was gently inserted through the anus and the faecal material present in the vicinity withdrawn. The gloves were replaced with every new evacuation and each faecal sample was placed into a sterile zip-lock collection bag.

Milk was aseptically collected as described by Pang *et al*., (2018) where the teat ends were strategically scrubbed clean for 10 to 15 sec with moist cotton balls, impregnated with 70% ethanol (EtOH). The first three streams of milk from each teat were discarded, thereafter composite samples (*i.e*. three streams of milk from each of the four teats) were collected into sterile sampling vials. The vials were securely capped and temporarily stored in a cooler box.

Blood samples from the cows were collected targeting the coccygeal vein as described by Shabbir *et al*., (2013). While still restrained, the tail of a cow was held vertically until it was horizontal to the

ground. The groove lying in the ventral midline of the tail was located and a site swabbed with cotton pads impregnated with 70% EtOH before venipuncture. Midway along the body of the coccygeal vertebra, a sterile 21 gauge, 25 mm sterile hypodermic needle was inserted perpendicularly to the surface of the skin to a depth of a few millimeters (to ensure minimal pain was inflicted to the animal) and blood withdrawn. The blood was collected into sterile 4 ml EDTA coated vacuum tubes. In order to ensure adequate hemostasis after removal of the needle, pressure was applied with gauze for 30 to 60 s.

Coincidentally, a cow with a retained placenta after spontaneously aborting was sampled. The placental sample was collected using a sterile no. 12 razor blade. Three pieces of tissue were cut from different places of the retained placenta, taking care to locate regions with cotyledons. The tissue snips were each about 2 - 5 centimeters in length and these were transferred into a sterile zip-lock collection bag.

All the collected samples were stored in properly labelled cooler boxes containing ice packs and transported to the Potchefstroom Provincial Veterinary Laboratory, North West Province, where they were initially aseptically processed to minimise contamination and thereafter stored at -20 °C until further analysis.

3.2.2. Genomic DNA extraction

Extraction of DNA on the pre-processed samples was performed at the Molecular Parasitology and Zoonosis Research Group laboratory at the Unit of Environmental Sciences and Management, North West University (NWU), Potchefstroom campus.

The Quick DNA Fecal/Soil Microbe extraction kit (Zymo Research, catalog no. D6010, Inqaba Biotechnical Industry, (Pty, Ltd.) was used to isolate microbial DNA from ≤ 150 mg of faecal samples. Microbial DNA from milk and blood samples (200 µl) was extracted using the Quick DNA Miniprep kit (Zymo Research, catalog no. D3024 & D3025, Inqaba Biotechnical Industry, (Pty) Ltd.), while catalog no. D4068 & D4069 was used to extract DNA from ≤25 mg of placental tissue. For optimal performance, beta-mercaptoethanol was added into to the Genomic Lysis Buffers to a final dilution of 0.5% (v/v) for all protocols.

For faecal DNA extraction, approximately 150 mg of faecal samples were added to the ZR BashingBead™ Lysis Tubes, followed by addition of 750 µL BashingBead™ Buffer. The tubes were then placed in a bead beater (TissueLyser LT, Qiagen) fitted with a 2 mL tube holder assembly and processed at maximum speed for 5 min. The ZR BashingBead™ Lysis Tubes were then centrifuged at 10,000 x g for 1 min in a microcentrifuge. Thereafter, 400 µL of the supernatants were transferred into Zymo-Spin™ III-F Filter held in Collection Tubes and centrifuged at 8,000 x g for 1 min. To each of the filtrates in the Collection Tubes, 1,200 µL of Genomic Lysis Buffer was added and the solutions mixed well with a micropipette. Aliquotes of 800 µL of the mixtures were transferred into Zymo-Spin™ IICR Columns held in Collection Tubes and centrifuged at 10,000 x g for 1 min. The flow through from the Collection Tubes were emptied and this step was repeated on the same Zymo-Spin™ IICR Columns and Collection Tubes. A total of 200 µL of DNA Pre-Wash Buffer was added to each of the Zymo-Spin™ IICR Columns in new Collection Tubes and centrifuged at 10,000 x g for 1 min. The g-DNA Wash Buffer (500 µL) was added to the Zymo-Spin™ IICR Columns and centrifuges at 10,000 x g for 1 min. The Zymo-Spin™ IICR Columns were then transferred into clean 1.5 mL microcentrifuge tubes and 100 µL (50 µL minimum) of DNA Elution Buffer was added directly to the column matrices of each tube. This was centrifuged at 10,000 x g for 30 s to elute the DNA. Then, Zymo-Spin™ III-HRC Filters were placed in clean Collection Tubes and 600 µL of Prep Solution added. These were centrifuged at 8,000 x g for 3 min. Finally, the eluted DNA were transferred to prepared Zymo-Spin™ III-HRC Filter tubes held in clean 1.5 mL microcentrifuge tubes and centrifuged at 16,000 x g for 3 min to filter the DNA.

For milk and blood microbial DNA extraction, 200 µL of sample were placed in 1.5 mL microcentrifuge tubes and 800 µL of Genomic Lysis Buffer added to each sample. The mixtures were vortexed for 4 - 6 s and allowed to stand for 5 - 10 min at room temperature. They were then transferred into Zymo-Spin™ IIC Columns held in Collection Tubes. These were then centrifuged at 10,000 x g for one min and the tubes with the flow through discarded. The Zymo-Spin™ IIC Columns were transferred into new Collection Tubes, then 200 µL of DNA Pre-Wash Buffer was added to the spin columns and centrifuged at 10,000 x g for one min. A total of 500 µL of g-DNA Wash Buffer was added to the spin columns and centrifuged at 10,000 x g for 1 min. The spin columns were transferred into clean 1.5. mL microcentrifuge tubes and 50 µL of DNA Elution Buffer was added to the spin columns. This was followed by 2 - 5 min of incubation at room temperature and then centrifugation at top speed for 30 s to elute the DNA.

For the placental tissue DNA extraction, the three pieces of tissue (≤ 25 mg) were placed in 1.5 mL microcentrifuge tubes and washed three times with 600 µL of phosphate buffered saline (PBS) then centrifuged at full speed for 1 min. After each wash the supernatant was discarded and replaced with

another 600 µL of PBS. After the 3rd wash the tissues were transferred into new microcentrifuge tubes and 95 µL of nuclease free water, 95 µL of Solid Tissue Buffer (Blue) as well as 10 µL of Proteinase K were added to each of the tubes. The contents were then mixed thoroughly by vortexing for 10 - 15 s and then incubated at 55 ºC for 3 hours until the tissues had solubilized. Thereafter the contents were mixed thoroughly and then 2 volumes of Genomic Binding Buffer were added to the contents. The mixtures were vortexed for 10 - 15 s. They were then transferred into Zymo-Spin™ IIC-XLR Columns held in Collection Tubes and centrifuged at 12,000 x g for 1 min, then the Collection Tubes with the flow through were discarded. The Zymo-Spin™ IIC Columns were transferred into new Collection Tubes, then 400 µL of DNA Pre-Wash Buffer was added to the spin columns and centrifuged at 12,000 x g for 1 min, thereafter the collection were tubes emptied. A total of 700 µL of g-DNA Wash Buffer was added to the spin columns and centrifuged at 12,000 x g for 1 min, the collection tubes were again emptied. Thereafter, 200 µL of g-DNA Wash Buffer was added to the spin columns and centrifuged at 12,000 *x g* for 1 min and this time the tubes with the flow through were discarded. The spin columns were transferred into clean 1.5 mL microcentrifuge tubes and 50 µL of DNA Elution Buffer was added directly on the matrix of the tubes. This was followed by 5 min of incubation at room temperature and then centrifugation at top speed for 1 min to elute the DNA.

All the eluted DNA samples were quantified using Qubit® Fluorometer 4.0 (Invitrogen, Thermo Fisher) and stored at -20 °C for downstream molecular application.

3.2.3. Amplification and sequencing of the V3-V4 hypervariable region

The amplification and sequencing workflow on the Illumina MiSeq is shown on Fig 3.3 and discussed in the sub-sections below in detail. Two no template controls (NTCs) *i.e.*, NSCF-neg1 and NSCF-neg2 consisting of PCR and sequencing laboratory reagents as well as nuclease free water in place of experimental DNA template were incorporated in the amplification and sequencing steps and processed alongside the experimental samples.

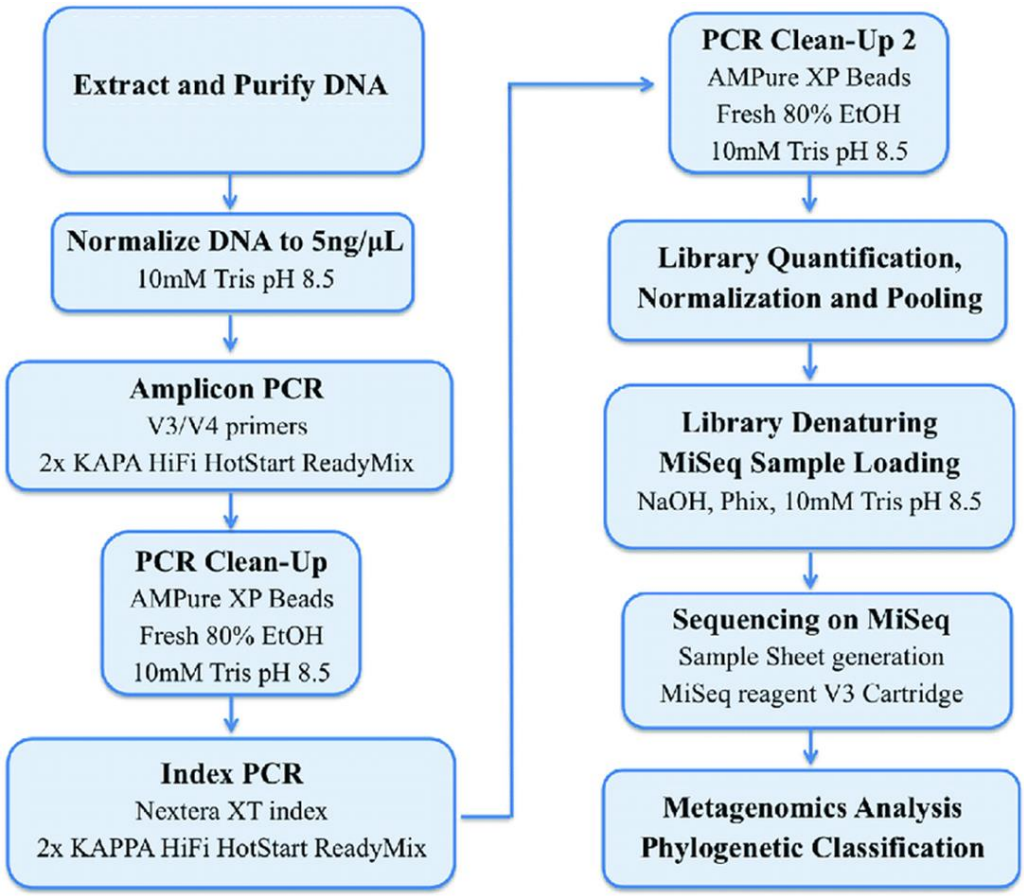

Fig 3.3: Workflow using the 16S Library Preparation Protocol in Illumina MiSeq platform, (Higuchi *et al*., 2018).

3.2.3.1. 16S rRNA library preparation

3.2.3.1.1. Amplicon PCR

The V3-V4 hypervariable region of the 16S rRNA was amplified with universal bacterial primers. The 16S Amplicon PCR Forward Primer: 5'TCGTCGGCAGCGTCAGATGTGTATA-AGAGACAGCCTACGGGNGGCWGCAG and 16S Amplicon PCR Reverse Primer: 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATC were used along with the Illumina overhang adapters Forward overhang: 5' CGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[locus- specific sequence] and Reverse overhang: 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[locus-specific sequence] (Falentin *et al*., 2016). For the amplicon PCR: primers; 2X KAPA HiFi HotStart Ready Mix (Kapa

Biosystems Inc, Roche); and microbial DNA template (5 ng/µL) were prepared. To prepare the PCR Master Mix 5 µL of 0.2 µM forward primer, 5 µL of 0.2 µM reverse primer and 12.5 µL of 2X KAPA HiFi HotStart Ready Mix were added into a 1.5 mL microcentrifuge tube. A volume of 22.5 µL of PCR Master Mix was aliquoted into each well of the 96-Well Skirted PCR Plate. Then 2.5 µL of DNA template was added to the wells and the plate sealed using MicroSeal 'A' adhesive film. The PCR reaction was performed in a thermal cycler (Bio-Rad C1000 Touch™ Thermal Cycler, Bio-Rad) programmed at 95 °C for 3 min, followed by 25 cycles of: 95 °C for 30 s; 55 °C for 30 s; 72 °C for 30 s ; 72 °C for 5 min and final hold at 4 °C.

3.2.3.1.2. PCR Clean-Up 1

A volume of 20 µL of the Amplicon PCR mixture was transferred to each well of the 0.8 mL storage plate, then 15 µL of well mixed AMPure XP beads (0.75x) was added. The AMPure beads were mixed with the PCR mixture by using the microplate thermoshaker for 2 min at 1,800 rpm. This was incubated at room temperature for 5 min, then the plate was placed onto a magnetic stand for 2 min. The supernatant (28 µL) was carefully removed and discarded while the plate was still on the magnetic stand. The beads were washed with 190 µL of 80% EtOH and the plate incubated at room temperature for 30 s. The supernatant (200 µL) was carefully removed and discarded. Then the EtOH washing and supernatant discarding steps were repeated. The beads were then left to air-dry for 10 -15 min with the plate still on the magnetic stand, then the residual EtOH was removed. The plate was removed from the magnetic stand, 25 µL of resuspension buffer was added to each well and mixed by using the microplate thermoshaker at 1,800 rpm for 2 min. This was incubated at room temperature for 2 min and thereafter the plate was placed on the magnetic stand for 2 min. With the plate still on the magnetic stand, 20 µL of the resuspensions were carefully transferred into a newly labelled storage plate. The PCR product (1 µL) was run on Agilent 4200 TapeStation System (Agilent Technologies, Germany) to verify fragment sizes.

3.2.3.1.3. Index PCR

A total volume of 50 µL of the Master Mix was made by adding; 5 µL of PCR product, 10 µL of Nextera DNA Flex Index Primer Set, 25 µL of 2X KAPA HiFi HotStart Ready Mix and 10 µL of PCR Grade Water. The plate was covered using MicroSeal 'A' adhesive film and the contents mixed by using

BioShake® microplate thermoshaker (Quantifoil Instrument, GmbH) at 1,800 rpm for 1 min, then centrifuged at 300 x g for 1 min. The PCR amplification was performed in a thermocycler at 95 °C for 3 min, followed by 8 cycles of: 95 °C for 30 s; 55 °C for 30 s; 72 °C for 30 s; 72 °C for 5 min and a holding temperature of 4 °C.

3.2.3.1.4. PCR Clean-Up 2

A volume of 45 µL of Amplicon PCR mixture was transferred to each well of the 0.8 mL storage plate, then 33.75 µL of well mixed AMPure XP beads (0.75x) were added. The AMPure beads were mixed with PCR mixture using the microplate thermoshaker for 2 min at 1,800 rpm. The mixture was incubated at room temperature for 5 min. The plate was placed onto a magnetic stand for 2 min. The supernatant (70 µL) was carefully removed and discarded while the plate was still on the magnetic stand. Keeping the plate on the magnetic stand, the beads were washed with 190 µL of 80% EtOH, then incubated at room temperature for 30 s. The supernatant (200 µL) was carefully removed and discarded, then the EtOH washing and discarding step was repeated. The beads were left to air-dry for 10 –15 min to remove residual EtOH while the plate was still on the magnetic stand. The plate was then removed from the magnetic stand and 30 µL of RSB was added to each well. This was mixed by using the microplate thermoshaker at 1,800 rpm for 2 min, incubated at room temperature for 2 min and the plate was placed on the magnetic stand for 2 min. A total volume of 25 µL of the resuspensions were carefully transferred to a newly labelled 0.8 mL storage plate whilst keeping the plate on the magnetic stand.

3.2.3.2. Library quantification, normalization and pooling

After purification of the PCR library, the next step included measuring the concentrations using Qubit®2.0 Fluorometer and verification of the size of the library PCR preparations by running 1 µL of the Index PCR product on the Agilent 4200 Tapestation system (Agilent Technologies, Germany).

The library was normalised to 2 nM concentrations using Dilution Buffer containing 10 mM Tris-HCl (pH 8.5) with 0.1% Tween 20 as determined by Agilent Technologies 2100 Bioanalyzer trace (Agilent Technologies, Waldbronn, Germany) using the formula below (Illumina, Inc):

$$\frac{\text{Concentration in ng/µl}}{660\ \text{g/mol} \times \text{average library size}} \times 10^6 = \text{Concentration in nM}$$

After normalising each library to 2 nM, the DNA libraries were pooled into one 1.5 mL microcentrifuge tube to create a pooled library. Thereafter the 5 µL pooled DNA samples were denatured using 5 µL of 0.2 N NaOH to create single strands ready for sequencing.

3.2.3.3. MiSeq sample loading

A volume 490 µL of pre-chilled Hybridization Buffer (HT1) was added to the tube containing 10 µL denatured DNA to make a final volume of 500 µL. This resulted in 20 pM of denatured libraries. Samples were briefly vortexed and centrifuged at 2 500 rpm for 1 min and incubated at room temperature for 5 min to allow all DNA to denature into single strands and the samples were placed on ice until final dilution stage.

3.2.3.4. Denaturation and dilution of PhiX control

The PhiX control was prepared by diluting 2 µL of 10 nM PhiX library into 3 µL of 10 mM Tris-Cl, pH 8.5 with 0.1% Tween 20, to obtain a concentration of 4 nM PhiX control. The PhiX control was then denatured by combining 5 µL of 4 nM PhiX library with 5 µL of 0.2 M NaOH to make a volume of 10 µL. The mixture was vortexed briefly and centrifuged at 280 x *g* for 1 min and incubated at room temperature for 5 min to denature the PhiX library. The denatured PhiX library was diluted to 20 pM by adding 10 µL of the denatured library to 990 µL of pre-chilled HT1 to make a volume of 1,000 µL, resulting in a concentration of 0.2 pM. The mixture was briefly vortexed and then pulse-centrifuged. The PhiX library was then stored at -20 °C for further use.

Subsequently, 60 µL of the PhiX control was added to 180 µL of the denatured and diluted library pool in a 1.5 mL microcentrifuge tube. A volume of 357.5 µL of pre-chilled HT1 and 2.5 µL of gBlocks Fragments were added to the tube. The microcentrifuge tube was briefly and gently vortexed, then centrifuged for 1 min. The solution was placed on ice, thereafter 600 µL of it were loaded onto the MiSeq flow cell and analysed using the MiSeq Controlled Software (MCS).

3.2.4. Metagenomic data and statistical analyses

3.2.4.1. Processing of 16S rRNA amplicon sequences

The generated raw sequences were recovered in FASTQ format from the Illumina MiSeq machine. These were initially quality checked using Sequencing Analysis Viewer (SAV) software compatible with MiSeq. Thereafter, FastQC (v0.11.8) for Microsoft Windows 2013 and trimGalore (v0.6.4_dev; https://github.com/FelixKrueger/TrimGalore) were used to determine the quality of the individual FASTQ reads (including sequence base content, Kmer frequency, GC content, sequence length, sequence duplication and adapter contamination) and filtering (adapter removal and read trimming, *i.e.* removing initial 13 base pairs from each individual FASTQ read). Only reads with the percentage of bases with a quality score of 20 or higher (Q ≥ 20) and length of at least 50 base pairs (≥ 50 bp) were considered for downstream analysis.

3.2.4.2. Assignment of Amplicon Sequencing Variants (ASVs), classification, abundance estimation and visualization

All the downstream analyses were performed in R (v3.6.1). Clean reads were pre-processed using the high-resolution Divisive Amplicon Denoising Algorithm 2 (DADA2) package (v1.12.1) including quality inspection, trimming, de-replication, merging paired-end reads, identification and removal of chimeric sequences as well as amplicon sequence variant (ASV) inference. Briefly, trimming and filtering was performed on Illumina-sequenced paired reads jointly after the filter pass. They were de-replicated to remove redundancy and the resulting de-replicated objects were assigned sample names. Pooled sequencing read inference was performed and the inferred forward and reverse sequences were matched removing paired sequences that did not perfectly overlap as a final control against residual error.

A sample-by-sequence feature table was constructed and chimeras were removed. Taxonomy was assigned to the obtained ASVs and the ASV abundance estimates determined using SILVA SSU taxonomic training data formatted for DADA2 (v138, 99% 16S full-length; McLaren 2020, https://zenodo.org/record/3986799#.YG7YR-gzY2w). ASVs assigned as Archaea and Eukaryota

were filtered out and further analyses were conducted only on Bacteria. To eliminate host DNA, ASVs were aligned against the Cow reference genome, *Bos taurus* ARS-UCD1.3 (https://www.ncbi.nlm.nih.gov/genome/?term=bos+taurus) and the matched ASVs were filtered out.

The sample-by-sequence feature table, the sample metadata and the sequence taxonomies were combined into a single PhyloSeq object saved as a CSV comma delimited file with columns for sample attributes such as sample ID, with the names of each of the samples and rows containing information on the sampling conditions related to each sample.

In PhyloSeq package (v1.28.0) as described by McMurdie and Holmes (2013), further filtering and denoising was conducted. Alpha (α)-diversity indices (Chao1 and ACE richness indices as well as Shannon and Simpson's diversity indices) were estimated using plot_richness function from PhyloSeq and mean comparison p-values calculated using stat_compare_means function from the ggpubr package (v0.4.0) and plotted with ggplot2 v3.2.1 (Wickham, 2016). Prior to calculating beta (β)-diversity, read counts were normalized (log2-fold-change) with DESeq2 (v1.24.0) as described in Love *et al*., (2014). Ordinations for β-diversity between samples was estimated using Principal Component Analysis (PCA) based on transformed data using Hellinger distance metric.The β-diversity between groups was estimated using Principle Coordinate Analysis (PCoA) based on weighted Unique Fraction (UniFrac) and Bray distance measurew; as well as Non-Metric Multidimensional Scaling (NMDS) based on Bray distance measure implemented in the plot_ordination and amp_ordinate functions in PhyloSeq package (v1.28.0) and the ampvis2 package (https://madsalbertsen.github.io/ampvis2/articles/ampvis2.html), respectively.

UpsetR v1.4.0 was used to construct intersection plots depicting the shared bacterial families and genus-level taxa between the different sample groups (Conway *et al*., 2017). Stacked bar plots of taxa present within and between samples at phylum and genus levels were plotted in Microsoft Excel (Windows 10). Differential abundance analysis was also performed using the negative binomial Wald test model implemented in DESeq2 as described in Love *et al*., (2014) on normalized data. After converting the PhyloSeq-format microbiome data into a *DESeqDataSet,* it was then ordered into a table by the adjusted *P*-value (*Padj*)/ q-value according to the ASVs that were among the most significantly differentially abundant between sample types and the results were plotted using some ggplot2 commands. The following pairwise combinations were used in the analysis: Blood vs Faeces, Blood vs Milk and Faeces vs Milk.

3.2.4.3. Statistical analysis

The significant differences in α-diversity were calculated between groups using Kruskal-Wallis test in PhyloSeq. Effect sizes of the differences between groups were calculated using the Cohen's D measure using the effsize package in R (https://github.com/mtorchiano/effsize), based on Shannon diversity indices. Significance for clustering on ordination plots was determined by Permutational Multivariate Analysis of Variance (PERMANOVA) using permutation test with pseudo *F* ratios as implemented in the Adonis function in the Vegan package (https://github.com/vegandevs/vegan).. Sample groups were used as independent variables and taxa prevalences at ASV (for α-diversity) and genus (for β-diversity) levels as dependant variables. *P*-values < 0,05 were deemed statistically significant.The ASVs were considered significantly differentially abundant (DA) when *Padj* was < 0.01 between sample types. Positive log2-fold change indicated increased abundance, while negative log2-fold change indicated decreased abundance.

3.2.5. Screening of selected pathogens

3.2.5.1. PCR detection of *Anaplasma* species

For detection and characterization of *Anaplasma* species, PCR was conducted using the 2X Kapa HiFi Hotstart ReadyMix. A total volume of 25 µl was prepared for the reaction in a 1.5 mL microcentrifuge tube constituted of 12.5 µl of Master Mix, 1 µl (2 µM concentration) of each of the forward primer and reverse primers (on Table 3.1), 3 µl of DNA template and a volume of 7.5 µl of nuclease free water to adjust the reaction volume.The *Anaplasma* positive controls were obtained from the University of Pretoria's Department of Veterinary Tropical Diseases (DVTD), in the Faculty of Veterinary Science. They consisted of *A. marginale* and *A. centrale* (vaccine strain) confirmed via Sanger sequencing in another study. The PCR conditions were set as follows on the thermal cycler: initial denaturation at 98°C for 10s, followed by 35 cycles of denaturation at 98°C for 1 s, annealing at 55°C for 5 s, extension at 72°C for 15 s and a final extension cycle of 72°C for 1 min. The PCR products were separated by gel electrophoresis on 1.5% agarose stained with ethidium bromide and size-fractionated using 100 bp or 1kb ladder as DNA size markers (Promega, USA) and subsequently photographed under UV light (EnduroTM GDS gel documentation system, Labnet International Inc.).

3.2.5.2. Culture and isolation of *Brucella*

Culture and isolation of *Brucella* was done in collaboration with DVTD at their Biosafety Level 2 Plus (BSL2 plus) facility using selective CITA medium (mCITA) as described previously (Ledwaba *et al*., 2020). Briefly, placental tissue samples were individually placed in labelled petri dishes where small pieces were cut with sterile blades. The cut pieces of samples were transferred into 2 ml tubes that contained 500µl of PBS and homogenized with a Precellys 24 tissue homogenizer (Bertin Technologies, Montigny-le-Bretonneux France), then stored in the refrigerator overnight. They were then cultured the following day on mCITA. The plates were incubated at 37 °C in 5 - 10% carbon dioxide and monitored every day from day 3 - 10. For milk culturing, a total volume of 200 µl of each milk sample was transferred into a 2 mL microcentrifuge tube and centrifuged at full speed for 1 min and thereafter the sediment was mixed with cream from the same sample then spread over the mCITA medium, left to air-dry at room temperature and incubated similarly to tissue samples. Any observed suspect colony was sub-cultured on mCITA as well as on Blood agar for clear visibility. A sample was considered negative if no *Brucella* colony was observed and isolated in the period of 10 days.

3.2.5.3. PCR detection of *Brucella* species

DNA from the positive cultures was extracted as described for milk and blood above and the identity of the isolates were confirmed with PCR. AMOS-PCR was conducted as described previously (Bricker & Halling, 1994 & 1995; Weiner *et al*., 2011) on DNA which was extracted directly from milk and placental samples as well as those grown on *Brucella* selective media using the Quick DNA Miniprep kit from Zymo Research (catalog no. D3024 & D3025, Inqaba Biotechnical Industry, (Pty) Ltd.). Amplification was conducted using the *OneTaq* Quick-Load 2X Master Mix (M0486L, New England BioLabs Inc). In addition to the Master Mix, the PCR mixture contained a combination of five primers specific for *B. abortus, B. melitensis, B. ovis, B. suis* (0.2 µM) and IS711 (1 µM), respectively (Table 3.1.) and 10 ng DNA per 25 µl reaction. The PCR conditions were optimized to an initial denaturation step of 95 °C for 3 min followed by 35 cycles of 95 °C for 1 min, annealing at 60 °C for 2 min and 72 °C for 2 min.

Positive controls for *Brucella* were also obtained from the DVTD. The *Brucella* controls consisted of Sanger sequencing confirmed DNA isolates of *B. canis* and *B. abortus* (S19 strain). PCR products were electrophoresed on a 1.5% agarose, size fractionated and photographed as described above.

3.2.5.4. DNA Purification

To purify the PCR products for sequencing, PureLink® Quick Gel Extraction Kit (LifeTechnologies, ThemoFischer) was used. Fragments of interest were excised from the gel under LED transilluminator (GeneDireX, Inc). The gel slices containing the DNA fragments were weighed using a scale sensitive to 0.001 g and dissolved in a Solubilization Buffer (L3) at a ratio of 3:1 buffer to weight of gel piece. The solubilisation occurred in a heat block set up to 50°C for 10 mins with gentle inversion of the 2 mL microcentrifuge tubes every 3 min to ensure gel dissolution. After the gel slices had dissolved the tubes were further incubated for 5 min. For optimal DNA yields, 1 gel volume of isopropanol was added to the dissolved gel slices. The DNA was purified following the microcentrifuge protocol. Prior to centrifugation, ethanol was added to the Wash Buffer (W1) according to the label on the bottle. Thereafter, the dissolved gels were loaded into Quick Gel Extraction Columns placed inside Wash Tubes. To bind the DNA, the columns were centrifuged at >12,000 × g for 1 min. The flow-through was discarded and the columns were placed back into the Wash Tubes. This was followed by a wash step which entailed adding 500 µL of W1 containing ethanol to the columns. The buffer was removed by centrifuging the columns at >12,000 × g for 1 min. The flow-through was discarded and the columns were placed into the Wash Tubes. To remove ethanol, the columns were centrifuged at maximum speed for 1 - 2 min and the flow-throughs discarded. To elute the purified DNA, the columns were placed into Recovery Tubes, 50 µL of Elution Buffer (E5) added to each column, incubated for 1 min at room temperature and thereafter centrifuged at >12,000 × g for 1 minute to collect the DNA.The elution tube containing the purified DNA was stored at −20°C until the samples were sequenced.

Table 3. 1: Sequences and characteristics of the primers used in the study

PCR assay name & target gene	Primer name	Primer sequence (5'- 3')	Annealing temp	Expected Fragment size	Reference
***16S* rRNA**	FD1	AGAGTTTGATCCTGGCTCAG	55 ˚C	1470 bp	Weisburg *et al*.,1991
	RP2	ACGGCTACCTTGTTACGACTT			
AMOS *IS711*	BA-F	GAC GAA CGG AAT TTT TCC AAT CCC	60 ˚C	498 bp	Bricker & Halling, (1994; 1995); Bricker *et al*., (2003), Weiner *et al*., (2011)
	BM-F	AAA TCG CGT CCT TGC TGG TCT GA		731 bp	
	BO-F	CGG GTT CTG GCA CCA TCG TCG	60 ˚C	976 bp	
	BS-F	GCG CGG TTT TCT GAA GGT GGT TCA GG		285 bp	
	IS711-R	TGC CGA TCA CTT AAG GGC CTT CAT			

CHAPTER 4: RESULTS

For metagenomics analysis, a total of thirty-three (n = 33) lactating cows owned by four different farmers (A-D) were sampled and designated sample ID numbers 1 to 33. Most of the animals were apparently healthy and no signs of udder inflammation were observed except from one animal which also had a retained placenta following a spontaneous abortion.

Initially, a total of n = 99 samples were obtained from the 33 animals, comprised of sets of three samples per cow, *i.e*. n = 33 faeces (assigned the label WF); n = 33 milk (WM) and n = 33 blood (WB) and coincidentally a placental tissue sample (harvested from three different regions of the placenta), assigned the label P. During the DNA extraction process, sample sets from two animals were excluded due to poor visual quality of the milk. Therefore, DNA samples extracted from faeces, milk and blood of 31 cows were further processed. After extraction of DNA, the samples were pooled for 16S rRNA metagenomic sequencing according to sample origin as shown on Table 4.1. After pooling, a total of 34 DNA samples were sequenced using the Illumina Miseq sequencing platform (*i.e.* n = 11 faeces; n = 11 milk; n = 11 blood and; n = 1 homogenized placental tissue). Post-sequencing, results of three DNA pool sets were excluded due to poor sequencing depth of the generated sequences. Eventually, sequence results from only 22 animals that passed quality filtering could be used in the structure and community analyses (*i.e.* n = 8 faeces; n = 8 milk; n = 8 blood and; n = 1 placenta). The microbial communities per sample pool were analysed within and between groups.

Table 4. 1: DNA pooling strategy for Illumina sequencing based on animal ID and origin

Farm	Animal ID per pool	Faecal sample pools	Milk sample pools	Blood sample pools	Placental sample
A	1	1WF	1WM	1WB	P
B	2,3,4	2WF	2WM	2WB	-
B	5,6,7	3WF	3WM	3WB	-
B	8,9,10	4WF	4WM	4WB	-
C	11,12,13	5WF	5WM	5WB	-
C	14,15,16	6WF	6WM	6WB	-
C	17,18,19	7WF	7WM	7WB	-
D	20,21,22	8WF	8WM	8WB	-

A-D = hidden identity of animal owners & sample origin, W = Waaihoek the location of the collection site, F = faeces, M = milk, B = blood & P = placenta.

Sequences generated from this study were deposited in the National Center for Biotechnology Information's Short Reads Archive, under BioProject number PRJNA777568, Accession numbers SRI168760 - SRI168784. See Table 4.2 for the read counts (number of reads retained across each step) tracked through the DADA2 pipeline including the ASV counts and richness per sample pool.

The generated sequences from the NTCs were compared to sequences of experimental samples in the retrospective assessment of contamination step. The NTCs contained sequences corresponding to seven microbial genera. These included *Escherichia/Shigella, Pseudomonas, Bacillus, Ralstonia, Blautia, Anaerobacillus* and *Lawsonella*. The sequences of three microbial genera (*Escherichia/Shigella, Pseudomonas* and *Bacillus*) were shared between the NTCs. These were also present in some of the experimental samples however, each microbial taxon is represented by multiple ASVs and those that matched the sequences within the NTCs were variably present and far fewer in number to constitute contamination. Therefore, a decision was taken to retain all sequences in the analysis as true biological signals. Of the four remaining genera, two were uniquely present in NSCF-neg1 and two also in NSCF-neg2, none of which could be detected in the experimental samples. The NTCs were subsequently removed and not included in the community analyses.

Table 4. 2: Number of reads retained per step tracked through DADA2 pipeline, ASV counts and taxa richness per samples

Sample Identity	Input	Filtered	DenoisedF	DenoisedR	Merged	Nonchim	ASV counts per sample	Taxa richness per sample	Genus-level taxa per sample
1WB	30831	23668	21193	20116	17993	14855	142	28	20
1WF	101970	89343	84384	84656	53802	35490	1665	86	56
1WM	31524	22757	19540	18817	13618	10470	190	83	57
2WB	37832	29795	26149	25061	18464	15676	566	92	61
2WF	120775	107047	101680	101732	69825	43711	1972	93	62
2WM	33836	24604	21643	20613	16216	12745	304	141	91
3WB	25144	18552	15794	14793	11600	9596	234	54	31
3WF	115418	101725	95413	96348	62787	40083	1836	90	56
3WM	23438	15587	13155	12425	7999	7350	50	29	18
4WB	18151	13594	10928	10384	7712	6461	129	22	17
4WF	127070	110075	104106	104681	69786	45412	1729	98	61
4WM	30551	23126	20748	19803	15735	13214	428	171	110
5WB	14985	2179	1508	1204	989	925	18	5	4
5WF	149339	130260	123486	124049	83710	53354	1977	94	61
5WM	52169	45439	43799	42993	41096	23979	900	227	150
6WB	62699	52625	48302	47520	36155	26674	1107	111	74
6WF	101937	88637	83147	84083	56819	37056	1003	31	21
6WM	68522	59153	54868	54273	46051	33966	870	331	207
7WB	22798	15199	12529	11540	7714	6727	154	28	21

7WF	157534	138463	132096	132977	95225	58202	1371	33	23
7WM	28214	22199	19025	18238	12422	11103	369	126	84
8 WB	25994	18472	15819	14644	11337	9445	179	23	17
8WF	118059	102889	96864	97427	64294	42087	1079	28	19
8WM	107650	94758	88672	88493	69243	43430	1653	189	125
P	130598	114575	113242	113716	110490	45147	271	17	16
NSCF-neg1	3223	2829	2824	2827	2824	1113	70	5	5
NSCF-neg2	2838	2482	2444	2434	2399	1038	85	5	5

W = Waaihoek the location of the collection site, F = faeces, M = milk, B = blood & P = placenta, NSCF-neg1 and -neg2 = no template controls. ASV, amplicon sequence variant

4.1. FAECAL MICROBIAL COMPOSITION

4.1.1. Sequence analysis

Analysis of n = 8 pooled faecal sample units yielded a total of 992 102 Illumina reads from the V3-V4 hypervariable region of the 16S rRNA gene. The total number of retained reads after filtering, denoising, merging of the forward and reverse reads and removal of chimeras including mean number ($\bar{x}$) and standard error of the mean (±SEM) were 355 395 (44 424 ± 2591). The number of reads per sample pool ranged from 35 490 to 58 202 [median 42 899; standard deviation (SD) 7327].

The overall number of ASVs detected in faeces was 4691, all assigned at least at phylum level. Approximately 64% of bacterial sequences found in faecal samples were assigned to genus level. Several genera had multiple ASVs associated with them and only one ASV could be resolved to species level. Collapsing of the ASVs yielded a minimum of 19 (8WF) and a maximum of 62 (2WF) genus level taxa that were detected per sample pool.

4.1.2. Alpha and beta diversity analysis

Alpha diversity was estimated with four indices *i.e.* the richness estimators, Chao1 and ACE as well as the diversity indices, Shannon and Simpson's using raw reads. The obtained values are summarized in Table 4.3 showing the minima, maxima, SD and $\bar{x}$ ± SEM for the processed faecal samples (n = 8). There was a high microbial richness and abundance within the faecal samples, furthermore the taxa were evenly distributed and with high diversity indices indicating a good diversity of the faecal microbial community.

Table 4. 3: Summary of alpha diversity estimates within faecal samples

Faecal samples (n=8)	**Raw Reads**	**Chao1**	**ACE**	**Shannon**	**Simpson's**
Minimum	15808	1209	1246	6.416	0.997
Maximum	18088	1710	1762	6.725	0.999
SD	785	154	157	0.088	0.0002
$\bar{x}$ ± SEM	30630 ± 277	1498 ± 54	1569 ± 56	6.6 ± 0.03	0.998 ± 7.25E-05

$\bar{x}$ = mean; SD = standard deviation; SEM = standard error of the mean.

The kingdom Bacteria (99%) dominated faeces with a small fraction of Archaea also detected (1.0%) across all faecal samples as shown on Fig 4.1. Hereafter the kingdom Archaea was filtered out from the analyses and only kingdom Bacteria was used in downstream analysis.

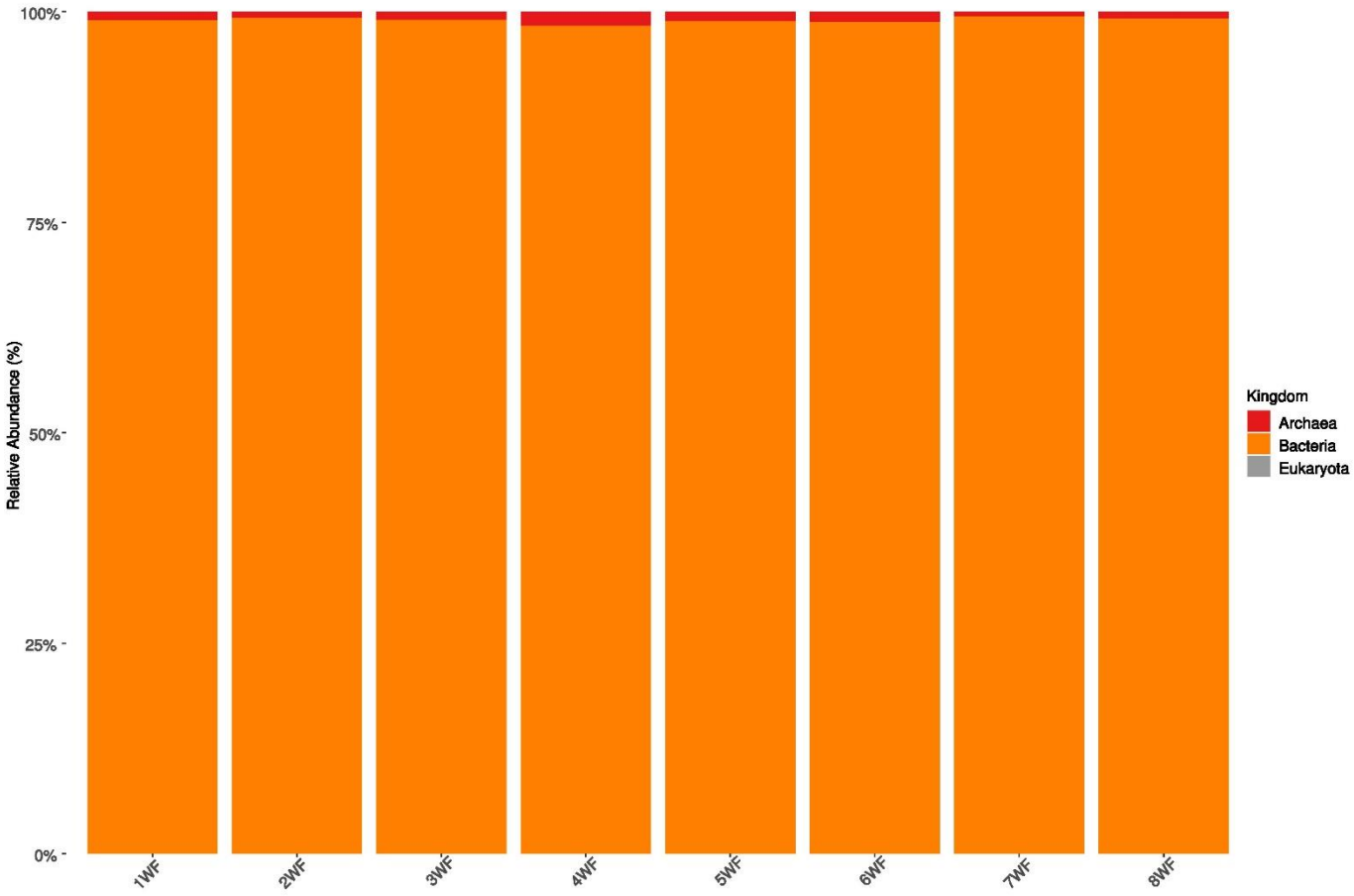

Fig 4.1: Relative abundance chart of the kingdom taxa (legend) detected in faecal samples (x-axis).

A total of 14 phyla were obtained in faeces with the top five most abundant being Firmicutes (64%); Bacteroidota (25.9%); Verrucomicrobiota (4.4%); Actinobacteriota (1.7%) and Patescibacteria (1.6%) accounting for 97.6% relative abundance of all assigned phyla in faecal samples. Firmicutes and Bacteroidota dominated faecal samples as expected (Fig 4.2).

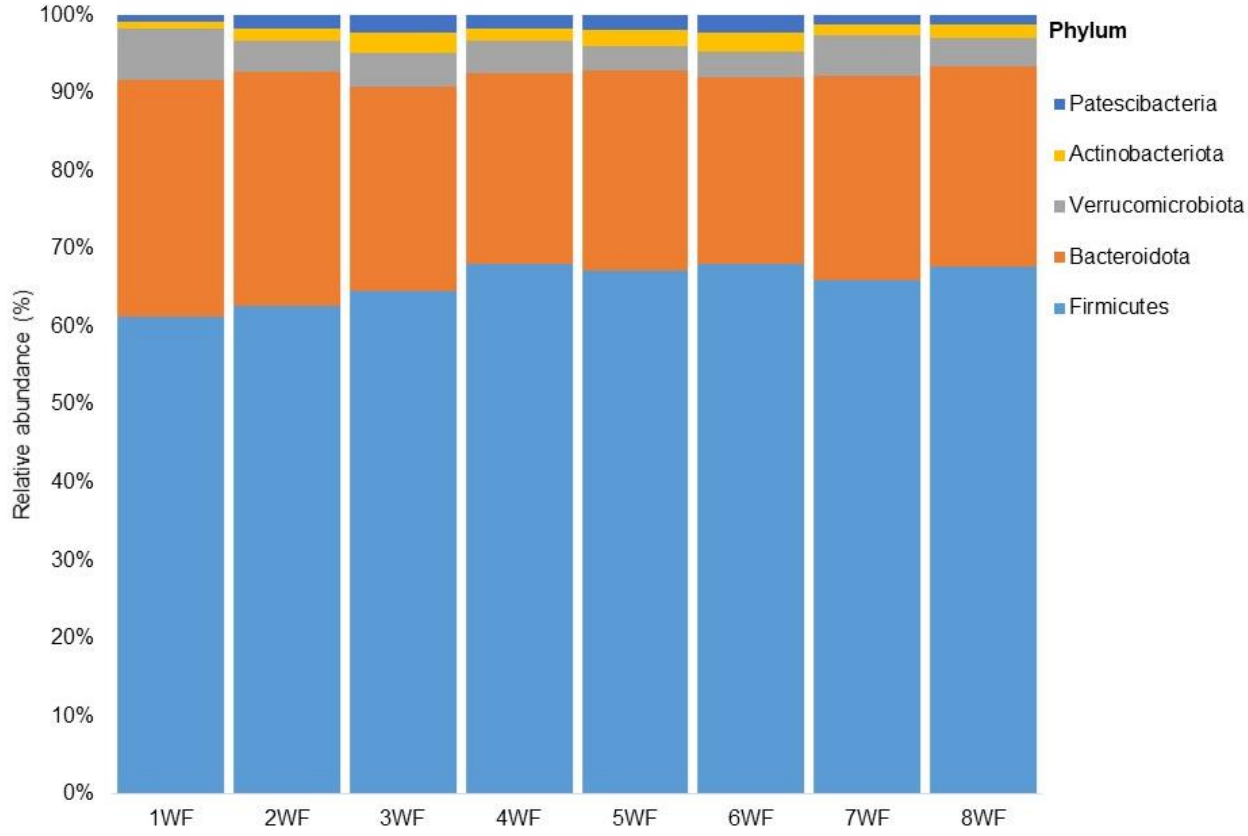

Fig 4.2: Relative abundance chart of the top 5 bacterial phyla (legend) in faecal samples (x-axis).

The top five most abundant classes out of a total of 21 obtained included Clostridia (60.9%); Bacteroidia (25.9%); Verrucomicrobiae (4.2%); Bacilli (3.0%) and Saccharimonadia (1.6%), which appeared under the following orders in descending order of abundance: Bacteroidales (26.2%); Peptostreptococcales-Tissierellales (24.5%); Oscillospirales (20.9%); Monoglobales (4.3%) and Verrucomicrobiales (4.3%). A total of 38 orders could be resolved.

There were 55 families in faeces and the most dominant included Peptostreptococcaceae (26.5%), Prevotellaceae (10.1%), Oscillospiraceae (9.3%), Rikenellaceae (9.0%) and UCG-010 (5.9%) among others, Fig 4.3.

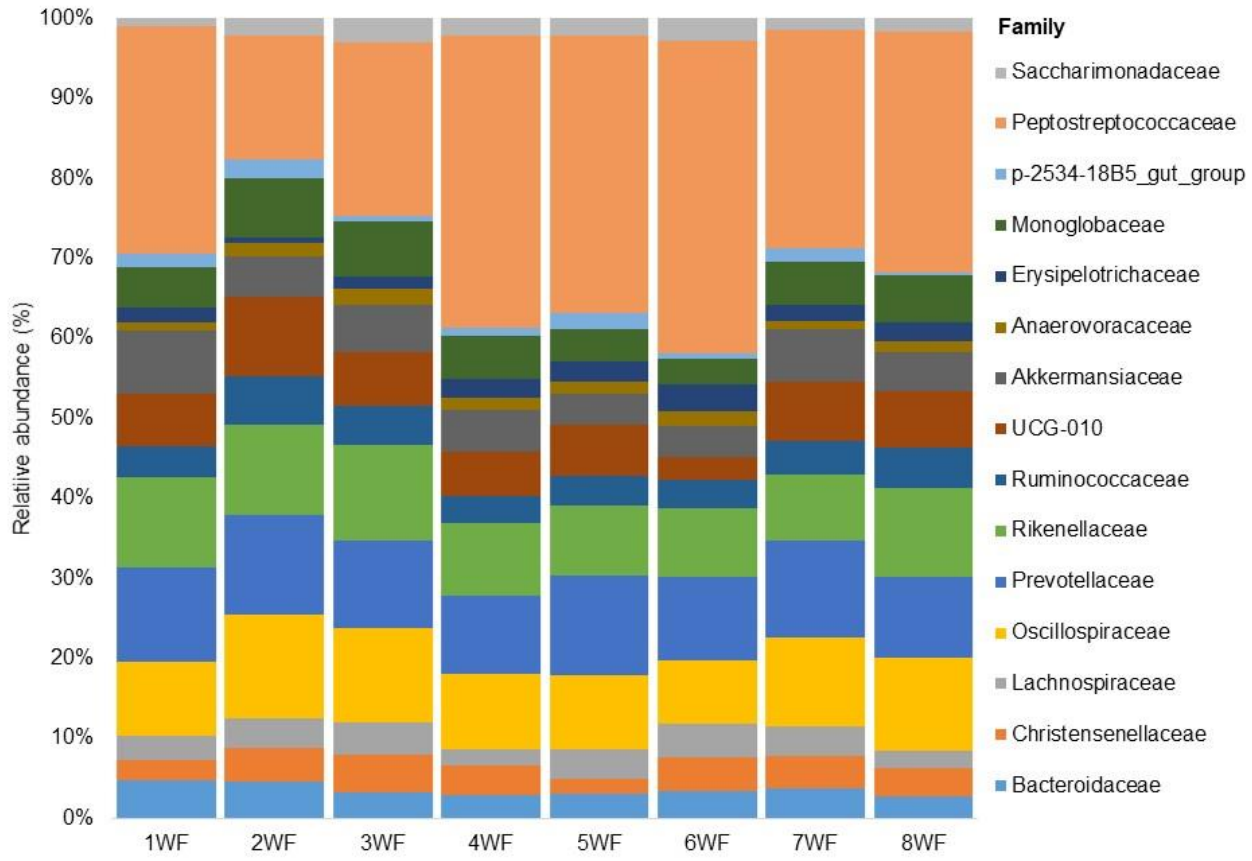

Fig 4.3: Relative abundance chart of the top 15 bacterial families (legend) in faecal samples (x-axis).

A total of 98 genus ranking taxa were detected among faecal samples, with the top 15 most abundant in descending order being *Romboutsia* (26.6%); *UCG-005* (8.9%); *Prevotellaceae_UCG-004* (8.3%); *Akkermansia* (6.0%); *Monoglobus* (6.0%); *Rikenellaceae_RC9_gut_group* (5.4%); *Bacteroides* (4.0%); *Christensenellaceae_R-7_group* (3.9%); *Paeniclostridium* (3.6%); *Alistipes* (3.4%); *Prevotellaceae_UCG-003* (2.8%); *Candidatus_Saccharimonas* (2.3%); *dgA-11_gut_group* (2.0%); *Turicibacter* (1.9%) and *Family_XIII_AD3011_group* (1.0%). These corresponded to 86.1% relative abundance of the assigned genus-level taxa contained in faeces with a distribution pattern greater than 1.0%. The microbial profile based on the most abundant taxa (15) present in 100% of the faecal sample pools was similar and was homogenous throughout, with minor variations in abundance as can be seen on the stacked bar plot (Fig 4.4) and accompanying Table 4.4 below. The dominant taxa were represented by members of Firmicutes (47%) and Bacteroidota (40%).

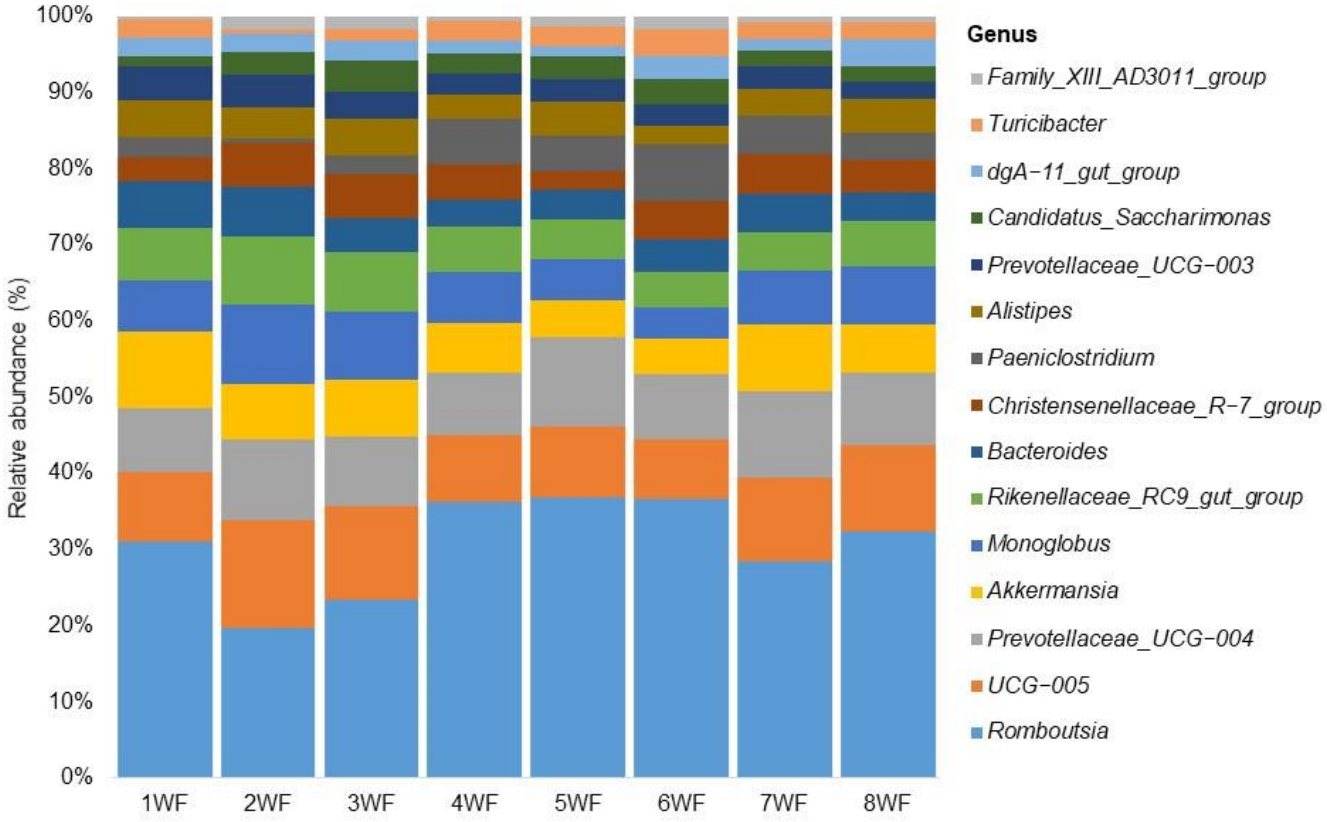

Fig 4.4: Stacked bar plot showing genus-level relative abundance of the top 15 taxa present in faecal samples.

Table 4. 4: Distribution of the 15 most abundant taxa at genus level across faecal samples

Genus	Group abundance	1WF	2WF	3WF	4WF	5WF	6WF	7WF	8WF
(F) *Romboutsia*	25687	3573	1857	2155	4053	3712	4388	2656	3293
(F) *UCG−005*	8553	1037	1337	1125	981	942	945	1030	1156
(B) *Prevotellaceae_UCG−004*	7959	961	1018	836	916	1175	1022	1064	967
(V) *Akkermansia*	5770	1164	673	691	727	503	563	805	644
(F) *Monoglobus*	5770	753	990	818	765	530	476	664	774
(B) *Rikenellaceae_RC9_gut_group*	5198	795	853	719	653	535	564	477	602
(B) *Bacteroides*	3863	709	**614**	**401**	398	399	506	462	374
(F) *Christensenellaceae_R−7_group*	3728	359	550	542	517	245	599	478	438
(F) *Paeniclostridium*	3437	308	**50**	**222**	665	470	891	472	359
(B) *Alistipes*	3289	542	388	435	371	452	304	336	461
(B) *Prevotellaceae_UCG−003*	2653	517	396	333	304	292	324	273	214
(Pa) *Candidatus_Saccharimonas*	2187	142	285	364	292	295	407	187	215
(B) *dgA−11_gut_group*	1956	285	223	243	193	144	356	151	361
(F) *Turicibacter*	1871	272	**68**	**137**	275	259	429	200	231
(F) *Family_XIII_AD3011_group*	974	61	159	166	82	138	206	83	79

Phylum indicated by letter(s) in parentheses before the taxon: F = Firmicutes, B = Bacteroidota, Pa = Patescibacteria, V = Verrucomicrobiota

The genus *Romboutsia* was the most abundant among faecal sample pools; recorded at 26.6% relative group abundance. A single ASV (ASV542) could be resolved to species level and it was a species within the genus *Romboutsia, i.e*. *R. sedimentorum,* present in all the faecal sample pools. Microbes of interest *i.e. Monoglobus* and *Akkermansia* were both detected across 100% of the faecal samples at a relative abundance of 6.0% each.

Among the faecal samples, genera of veterinary significance such as *Bacteroides* (4.0%); *Bacillus* (0.2%); *Prevotella* (0.1%); *Streptococcus* (0.1%); *Pseudomonas* (< 0.1%) and *Fusobacterium* (< 0.1%) were observed.*Bacteroides* and *Bacillus* were detected in all (100%) the faecal samples. *Prevotella* was detected in 5 (62.5%) of the samples, missing in sample 3WF, 6WF and 8WF. *Streptococcus* was detected in only three (37.5%) of the faecal samples *i.e*. 2WF, 4WF and 6WF. The genera *Pseudomonas* and *Fusobacterium* were detected in one sample each *i.e.* 2WF and 1WF, respectively.

Principal component analysis (PCA) using genus-level taxonomic ranks was performed to assess the β-diversity of the bacterial populations present between faecal samples. β -diversity was plotted using Hellinger distance metric (Fig 4.5). The analysis preserves variance of the samples and combines abundance and phylogeny to create contrasts that are used as input variables in the comparison of samples. The majority of the samples formed a cluster with little variation on the 1st and 2nd principal components. The first principal component accounted for 18% of the variation. The clustering of the samples was consistent with their similarity observed on the taxonomic bar charts.

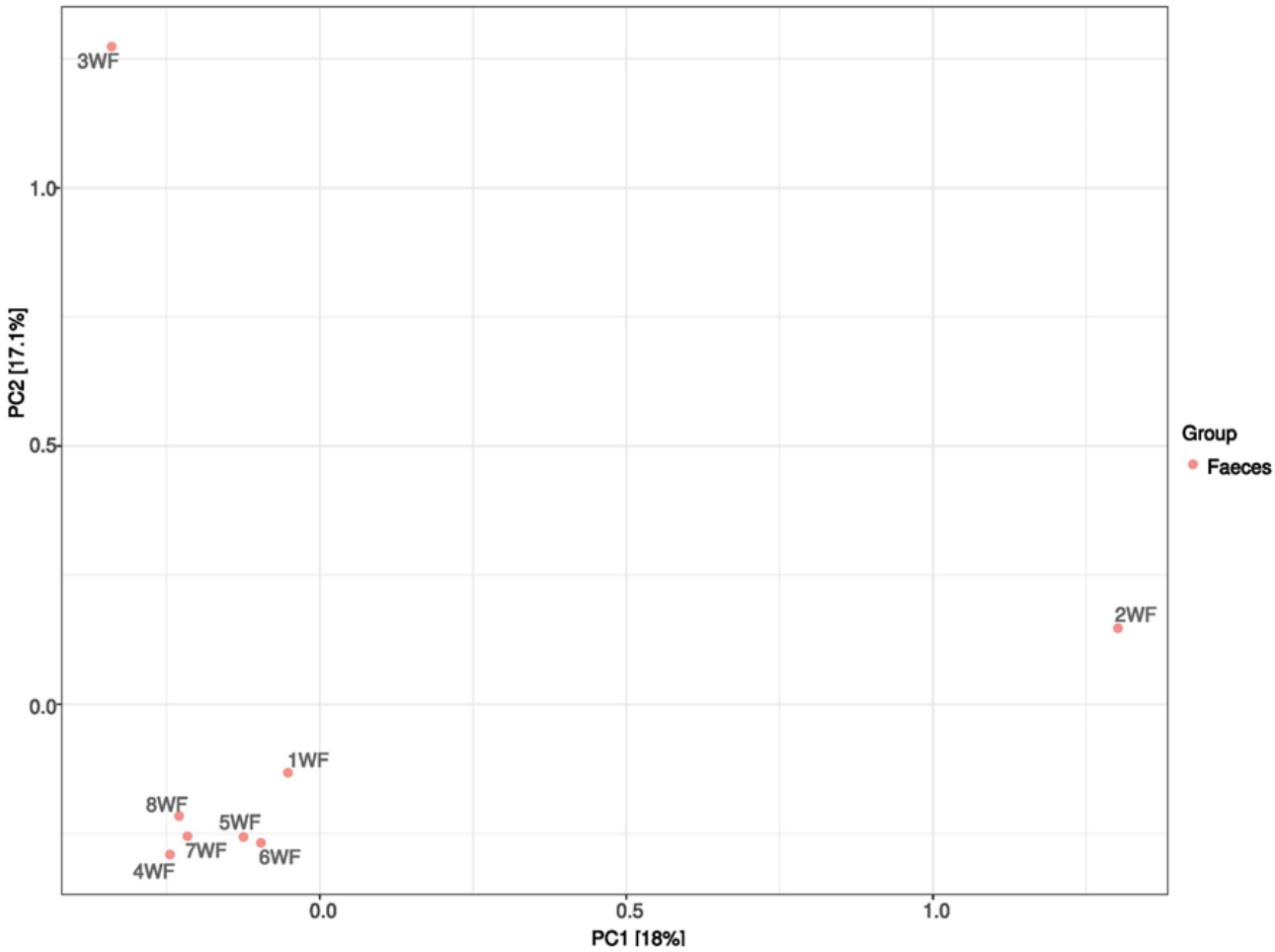

Fig 4.5: β-diversity of faecal samples shown with PCA using Hellinger distance metric.

This close clustering was expected having observed that the faecal samples generally possessed similar microbial communities with regards to the most abundant taxa as could be seen in Fig 4.4. Only two samples deviated from the cluster *i.e*. 2WF and 3WF respectively. Despite having similar communities it was observed that these two samples had the least abundance of the most dominant taxon *i.e*. *Romboutsia*, while also having the highest abundance of *Monoglobus* in comparison to the other faecal samples which could explain the deviation from the cluster (highlighted in red on Table 4.4). There were other subtle differences between sample 2WF and 3WF with regards to their composition based on the most abundant taxa (these are indicated in bold on Table 4.4) which could explain their divergence from each other, however not dismissing other unapparent possible contributors to the divergence. Further investigation into these two samples showed that sample 3WF had a high number of unique taxa (well over 120, present in lower abundances) that were absent in other faecal samples. On the other hand, further investigation into sample 2WF did not yield tangible information.

Overall there were no stark differences between faecal samples and the abundances were agreeing all across samples with minor observed differences between all the samples.

The core microbial taxa found in bovine faeces were catalogued as genus ranks that were consistently present among ≥ 75% of all samples with an overall relative group abundance of ≥ 0.1%. This microbiota consisted of 43 taxa listed in descending order of abundance on Table 4.5. This core taxa accounted for 96.8% of the taxa obtained in faeces. Among these, 32 taxa were present across all (100%) faecal samples, ranging between 0.1 - 26.6% in relative group abundance.

Table 4. 5: Core microbiota present in ≥ 75% of faecal samples at ≥ 0.1% group abundance including average and range across cow samples

Core microbiota	**Number of sequences per genus**	**Relative abundance (%)**	**Number of samples positive out of 8 (%)**	**Average & (range) across faecal samples (%)**
Romboutsia	25687	26.6	8 (100)	18.6 (10.7 – 24.3)
UCG-005	8553	8.9	8 (100)	6.2 (5.2 – 7.7)
Prevotellaceae_UCG-004	7959	8.3	8 (100)	5.8 (5.0 – 7.2)
Akkermansia	5770	6.0	8 (100)	4,2 (3.1 – 6.4)
Monoglobus	5770	6.0	8 (100)	4.2 (2.6 - 5,7)
Rikenellaceae_RC9_gut_group	5198	5.4	8 (100)	3.8 (3.0 – 4.9)
Bacteroides	3863	4.0	8 (100)	2.8 (2.1 – 3.9)
Christensenellaceae_R-7_group	3728	3.9	8 (100)	2.7 (1.5 – 3.3)
Paeniclostridium	3437	3.6	8 (100)	2.5 (0.3 – 4.9)
Alistipes	3289	3.4	8 (100)	2.4 (1.7 – 3.0)
Prevotellaceae_UCG-003	2653	2.8	8 (100)	1.9 (1.2 – 2.9)
Candidatus_Saccharimonas	2187	2.3	8 (100)	1.6 (0.8 – 2.3)
dgA-11_gut_group	1956	2.0	8 (100)	1.4 (0.9 – 2.0)
Turicibacter	1871	1.9	8 (100)	1.3 (0.4 – 2.4)
Family_XIII_AD3011_group	974	1.0	8 (100)	0.7 (0.3 – 1.1)
NK4A214_group	956	1.0	8 (100)	0.7 (0.5 – 0.8)
Clostridium_sensu_stricto_1	905	0.9	8 (100)	0,7 (0.1 – 1.5)
Treponema	766	0.8	8 (100)	0.6 (0.3 – 0.9)
DNF00809	718	0.7	8 (100)	0.5 (0.4 – 1.1)
Olsenella	653	0.7	8 (100)	0.5 (0.2 – 0.8)

Candidatus_Soleaferrea	624	0.6	8 (100)	0.5 (0.2 – 0.7)
Prevotellaceae_UCG-001	608	0.6	8 (100)	0.4 (0.2 – 0.6)
Coprococcus	606	0.6	8 (100)	0.4 (0.0 – 0.8)
Alloprevotella	577	0.6	8 (100)	0.4 (0.2 – 0.7)
p-1088-a5_gut_group	576	0.6	8 (100)	0.4 (0.2 – 0.7)
UCG-009	499	0.5	8 (100)	0.4 (0.1 – 0.6)
UCG-002	496	0.5	8 (100)	0.4 (0.2 – 0.5)
Odoribacter	423	0.4	8 (100)	0.3 (0.2 – 0.4)
Rikenella	258	0.3	8 (100)	0.2 (0.1 – 0.3)
Bacillus	213	0.2	8 (100)	0.2 (0.1 – 0.2)
Terrisporobacter	206	0.2	8 (100)	0.1 (0.0 – 0.2)
Agathobacter	179	0.2	7 (87.5)	0.1 (0.0 – 0.2)
Oscillibacter	177	0.2	7 (87.5)	0.1 (0.0 – 0.4)
GCA-900066575	145	0.2	7 (87.5)	0.1 (0.0 – 0.2)
Pseudoflavonifractor	143	0.1	6 (75)	0.1 (0.0 – 0.2)
Incertae_Sedis	141	0.1	8 (100)	0.1 (0.03 – 0.2)
Flexilinea	136	0.1	6 (75)	0.1 (0.0 – 0.2)
Dorea	133	0.1	6 (75)	0.1 (0.0 – 0.2)
Mogibacterium	117	0.1	7 (87.5)	0.1 (0.0 – 0.2)
Solobacterium	113	0.1	6 (75)	0.1 (0.0 - 0,2)
Saccharofermentans	106	0.1	6 (75)	0.1 (0.0 – 0.2)
Erysipelotrichaceae_UCG-009	89	0.1	6 (75)	0.1 (0.0 – 0.1)
Papillibacter	88	0.1	7 (87.5)	0.1 (0.0 – 0.1)

Genus-level taxa forming the core faecal microbiota in descending order of abundance of sequences and relative abundance. The number of cows that were positive for each genus, average and the range of the total bacterial population represented by each genus across all cows sampled is also shown in the table.

4.2. MILK MICROBIAL COMPOSITION

4.2.1. Sequence analysis

The pooled milk samples (n = 8) yielded 375 904 Illumina reads. The total number of non-chimeric reads obtained including $\bar{x}$ ± SEM were 156 257 (19 532 ± 4306). The number of reads per sample pool ranged from 7350 to 43 430 (median 12 980; SD 12 180).

The overall number of ASVs detected in milk was 4076. Approximately 67% of the sequences found in milk samples were assigned to genus level taxonomy. Several genera had multiple ASVs associated with them and only two ASVs could be resolved to species level. Collapsing of the ASVs yielded a minimum of 18 (3WM) and a maximum of 207 (6WM) genus level taxa that were detected per sample pool.

4.2.2. Alpha and beta diversity analysis

Alpha diversity of milk samples was estimated with four indices *i.e.* Chao1, ACE, Shannon and Simpson's, similarly to faecal samples. The obtained values are summarized in Table 4.6 showing the minima, maxima, SD and $\bar{x}$ ± SEM for the processed milk samples (n = 8). The milk samples had a variably rich microbial composition with extremes observed between the highest and lowest richness values. High diversity index values were recorded for milk samples, showing great diversity of taxa within the samples.

Table 4. 6: Summary of alpha diversity estimates within milk samples

Milk samples (n=8)	**Raw Reads**	**Chao1**	**ACE**	**Shannon**	**Simpson's**
Minimum	3024	50	50	3.638	0.9687
Maximum	17171	1444	1490	6.666	0.998
SD	4445	429	442	0.091	0.009
$\bar{x}$ ± SEM	12830 ± 1572	557 ± 152	563 ± 156	5.629 ± 0.320	0.993 ± 0.003

$\bar{x}$ = mean; SD = standard deviation; SEM = standard error of the mean.

The microbial communities in milk samples belonged to three kingdoms including Bacteria (75.6%), Eukaryota (23.2%) and Archaea (1.2%), shown on Fig 4.6. Eukaryota and Archaea were filtered out for subsequent downstream analyses.

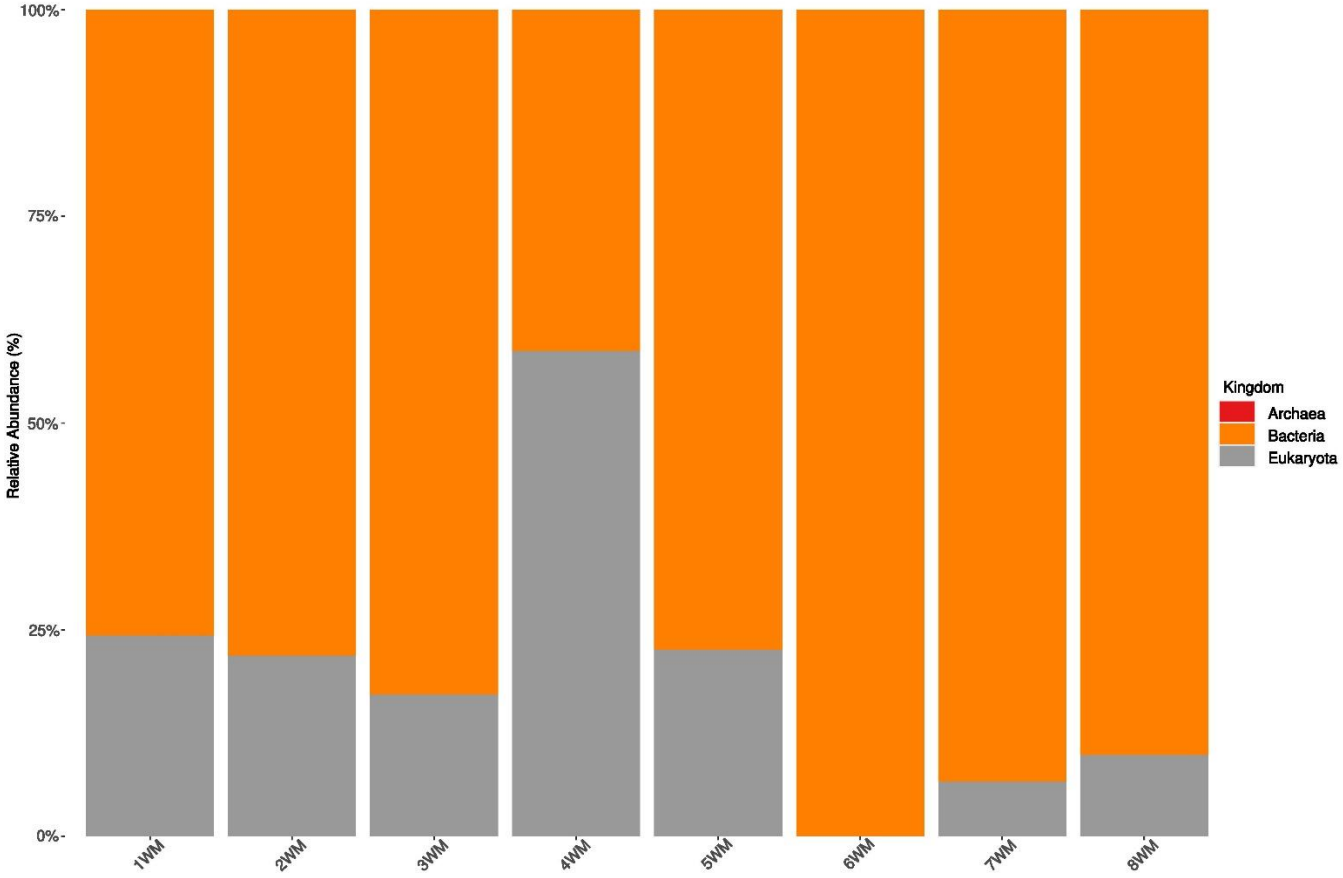

Fig 4.6: Relative abundance chart of the kingdom taxa (legend) detected in milk samples (x-axis).

A total of 31 phyla were obtained in milk samples with the top five most abundant in descending order of abundance being Firmicutes (39.4%); Bacteroidota (20.4%); Proteobacteria (15.0 %); Actinobacteriota (7.3%) and Verrucomicrobiota (3.2%), accounting for 85.3% relative abundance of all assigned phyla in milk Fig 4.7. Firmicutes and Bacteroidota, similarly to faeces dominated milk samples.

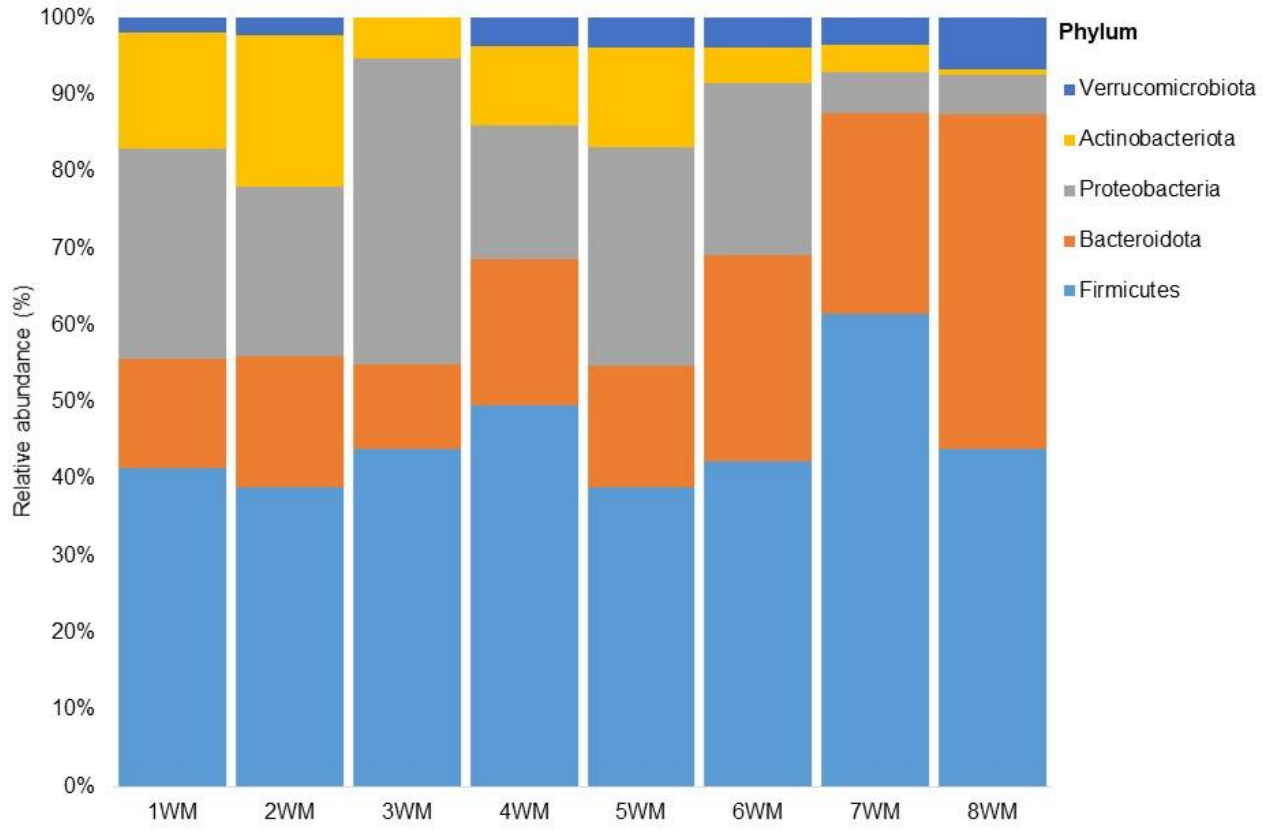

Fig 4.7: Relative abundance chart of the top 5 bacterial phyla (legend) in milk samples (x-axis).

There was a total of 74 classes, the top five most abundant included Clostridia (34.3%); Bacteroidia (20.0%); Gammaproteobacteria (9.0%); Alphaproteobacteria (6.0%) and Actinobacteria (4.7%).

The taxa grouped under 154 orders with the top five most abundant being Bacteroidales (18.8%); Oscillospirales (18.6%); Peptostreptococcales-Tissierellales (4.7%); Lachnospirales (3.9%) and Pseudomonadales (3.2%) in descending order of abundance.

The milk microbiota clustered under 236 families in total and the top five most abundant included Oscillospiraceae (8.8%); UCG-010 (7.6%); Rikenellaceae (5.6%); Lachnospiraceae (4.2%); and Prevotellaceae (3.8%). The distribution of the 15 most abundant families across the milk sample pools is shown on Fig 4.8.

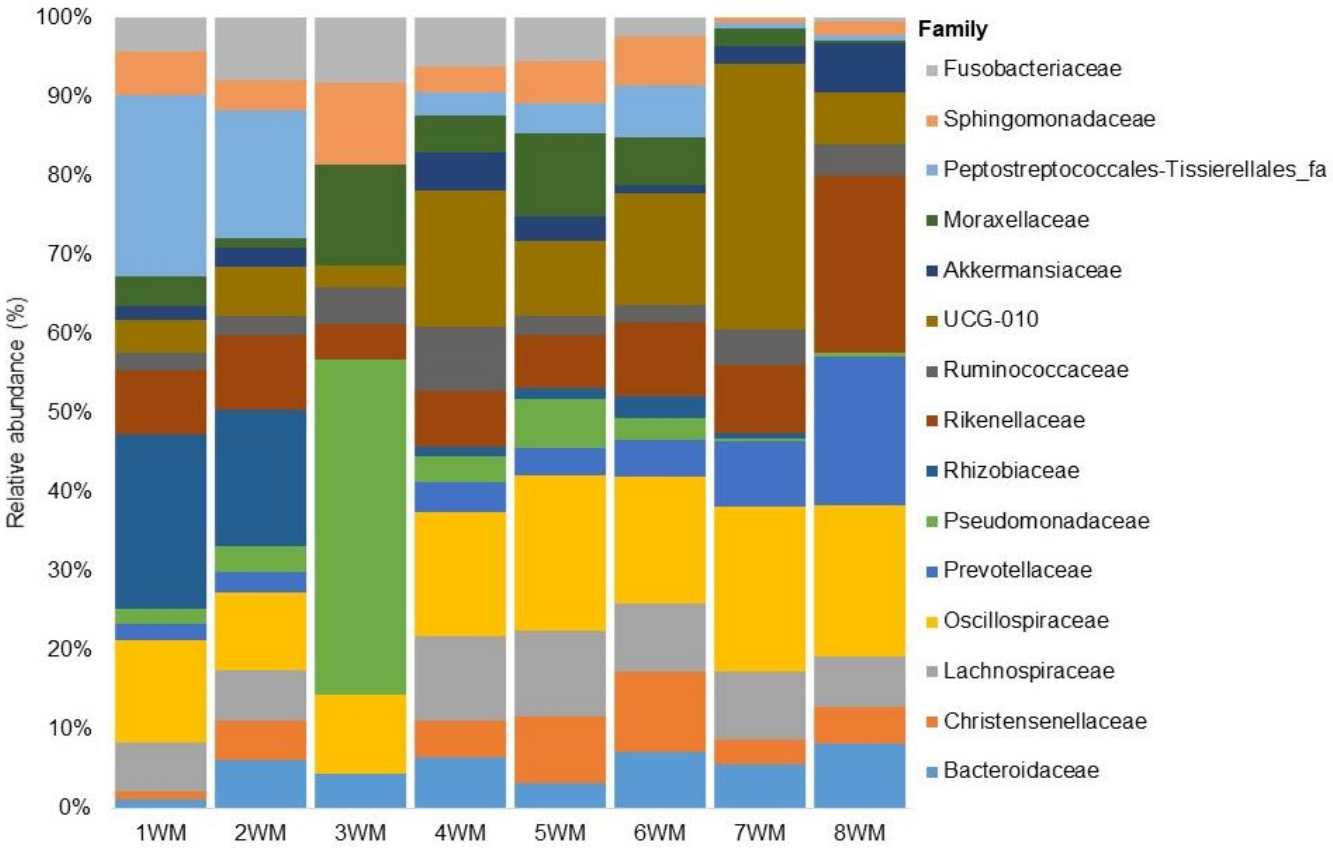

Fig 4.8: Relative abundance chart of the top 15 bacterial families (legend) in milk samples (x-axis).

A total of 374 genus ranking taxa were detected (Fig 4.9) from the milk samples and the top 15 most abundant in descending order of abundance were as follows: *UCG-005* (7.6%); *Bacteroides* (4.1%); *Rikenellaceae_RC9_gut_group* (4.0%), *Christensenellaceae_R-7_group* (3.3%); *Prevotellaceae_UCG-004* (3.1%); *Helcococcus* (3.0%); *Brucella* (2.6%); *Alistipes* (2.4%); *Akkermansia* (2.4%); *Fusobacterium* (2.3%); *Pseudomonas* (2.2%); *Monoglobus* (2.1%); *Rhodococcus* (1.5%); *NK4A214_group* (1.4%) and *Prevotellaceae_UCG-003* (1.2%). The most dominant microbial taxa (represented by members of Firmicutes; 33% and Bacteroidota; 33%) were variably present in unequal proportions across the milk samples as can be seen on Fig 4.9 and Table 4.7. These corresponded to 43.2% of the assigned genus-level taxa contained in milk samples.

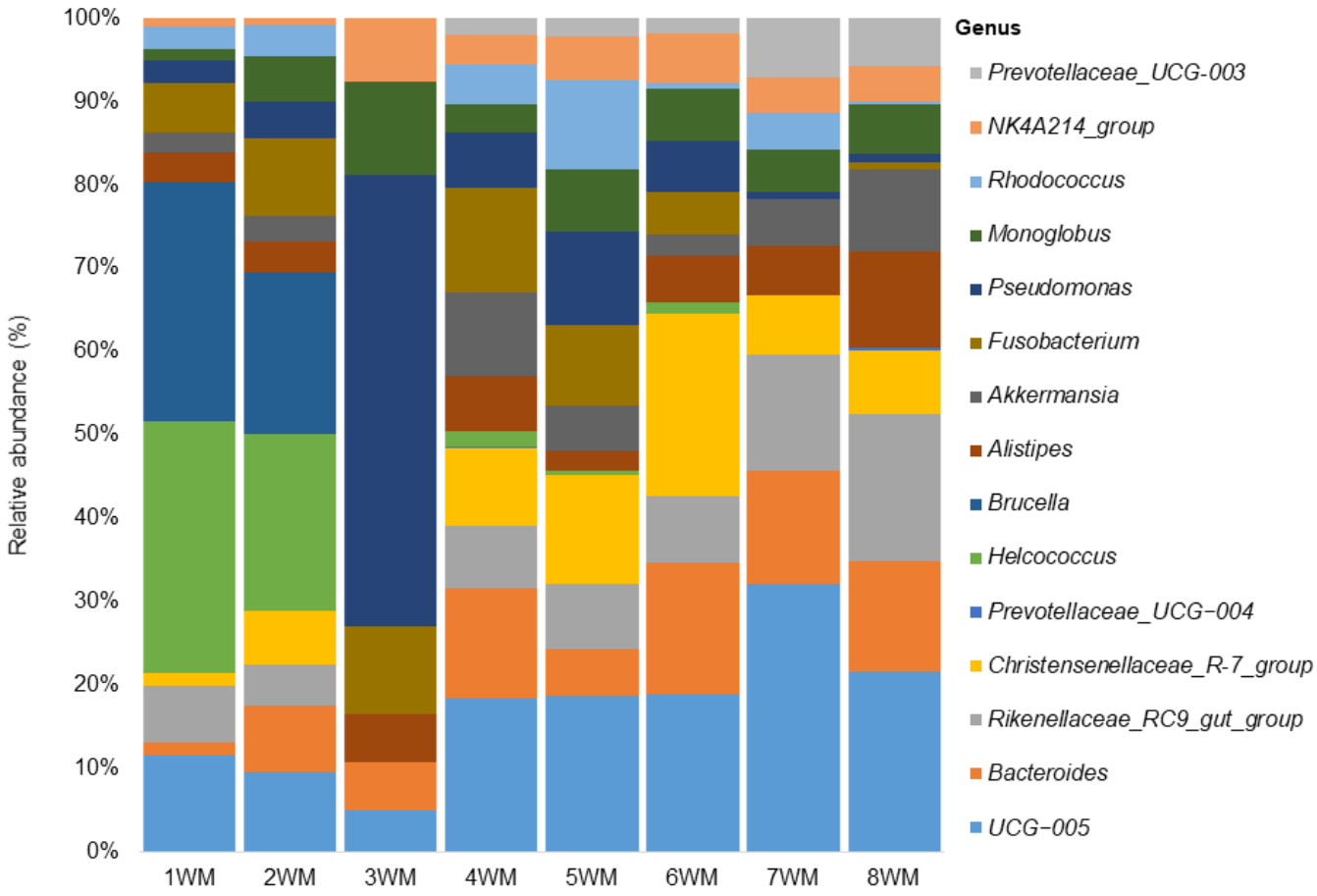

Fig 4.9: Stacked bar plot showing genus-level relative abundance of the top 15 taxa present in milk samples

Table 4. 7: Distribution of the top 15 most abundant taxa at genus level across milk samples

Genus	group abundance	1WM	2WM	3WM	4WM	5WM	6WM	7WM	8WM
(F) *UCG−005*	4948	375	390	38	608	590	346	1389	1212
(B) *Bacteroides*	2652	51	326	43	439	179	285	588	741
(B) *Rikenellaceae_RC9_gut_group*	2638	220	203	0	250	242	146	594	983
(F) *Christensenellaceae_R-7_group*	2163	49	259	0	309	415	395	308	428
(B) *Prevotellaceae_UCG−004*	2049	89	109	0	181	48	140	322	1160
(F) *Helcococcus*	1945	978	860	0	66	16	25	**0**	**0**
(P) *Brucella*	1709	925	784	**0**	**0**	**0**	**0**	**0**	**0**
(B) *Alistipes*	1593	114	152	43	215	71	102	253	643
(V) *Akkermansia*	1551	80	127	0	331	171	44	240	558
(Fu) *Fusobacterium*	1512	191	379	78	416	308	93	0	47
(P) *Pseudomonas*	1438	85	176	404	220	351	112	41	49
(F) *Monoglobus*	1613	44	222	84	112	236	113	216	335
(A) *Rhodococcus*	956	89	154	0	155	333	12	193	20
(F) *NK4A214_group*	934	34	34	57	122	167	108	180	232
(B) *Prevotellaceae_UCG-003*	805	0	0	0	65	69	35	310	326

Phylum indicated by letter(s) in parentheses before the taxon: A = Actinobacteriota; B = Bacteroidota, F = Firmicutes, Fu = Fusobacteriota, P = Proteobacteria, V = Verrucomicrobiota.

What could be observed from Table 4.7, is that sample 1WM and 2WM had high counts of *Brucella* and were the only two of the eight sample pools that contained this bacterium.

Sample 3WM lacked the majority of the dominant taxa, possessing only 7 of the 15 *i.e*. *UCG-005, Bacteroides, Alistipes, Fusobacterium, Pseudomonas, Monoglobus* and *NK4A214_group*.

In addition to *Brucella* (2.6%), taxa associated with conditions and/or diseases of veterinary significance that were identified include *Bacteroides* (4.1%); *Helcococcus* (3.0%); *Fusobacterium* (2.3%)*; Pseudomonas* (2.2%); *Rhodococcus* (1.5%); *Trueperella* (1.2%); *Porphyromonas* (1.1%); *Escherichia*/*Shigella* (1.0%); *Streptococcus* (0.9%); *Bacillus* (0.6 %); *Staphylococcus* (0.3%) and at <0.1% relative abundance *Mycobacterium; Legionella; Klebsiella* as well as *Mycoplasma*.

Hundred percent of the milk samples contained *Pseudomonas* and *Porphyromonas*. *Akkermansia* and *Rhodococcus* were detected in 7 out of 8 samples (87.5%). Most (75%) of the samples except for 2WM and 4WM contained *Escherichia/Shigella*. *Bacillus* was present in 5 (62.5%) samples (1WM and 3 - 7WM); similarly, *Streptococcus* (1WM & 4WM-7WM); while *Staphylococcus* was detected in four (50%) samples (2WM-3WM & 6WM-7WM). *Brucella* and *Trueperella* were only contained in sample pools 1WM and 2WM. *Helcococcus* was in high abundance in sample 1WM and 2WM, but was also detected in samples 4-6WM. *Mycoplasma* was detected in one sample (7WM); similarly, *Mycobacterium* (4WM); *Legionella* and *Klebsiella* (both in sample 6WM). Only two ASVs could be resolved to species level among milk samples *i.e*. *Fusobacterium necrophorum* (ASV149)*,* and *Luteimonas composti* (ASV13364). *F. necrophorum* was detected in four milk sample pools *i.e*. 2WM and 4-6WM; while *L. composti* could be detected in one sample pool *i.e*. 5WM.

To compare whole microbial composition within milk samples, PCA was conducted using genus-level taxonomic ranks. As shown in Fig 4.10, the clustering of the milk samples made it possible to separate them according to similarities in their microbial constituents, consistent with the stacked bar plot distribution pattern (Fig 4.9). Sample 1WM and 2WM grouped together, important to note is that both contained high ASV counts of the genus *Brucella* which was absent in all other samples (bold text on Table 4.7). These two samples also exclusively shared *Trueperella*, *Dietzia*, *Facklamia* and a taxon in the order Bacteroidales that was not resolved to genus level. Furthermore, they both contained higher proportions of *Helcococcus, Porphyromonas, Sphingomonas, Narcoidioides, Atopobium, Quadrisphaera* and *Flavobacterium* in comparison to the *Brucella* negative samples.

Sample 3WM was noisy from the rest of the cluster, upon further investigation it was further noted that this sample pool lacked 8 of the 15 most abundant taxa as discussed above and shown on Table 4.7 (red highlights), it further contained the least number of taxa and abundance overall in comparison to other milk samples.

The rest of the sample pools somewhat grouped close to one another. An observation was made that they possessed majority of the 15 most abundant taxa, mostly lacking *Brucella*. In addition to lacking *Brucella,* sample 7WM and 8WM also lacked the genus *Helcococcus*.

The samples were dispersed along the first two principal components with 32.1% total variation.

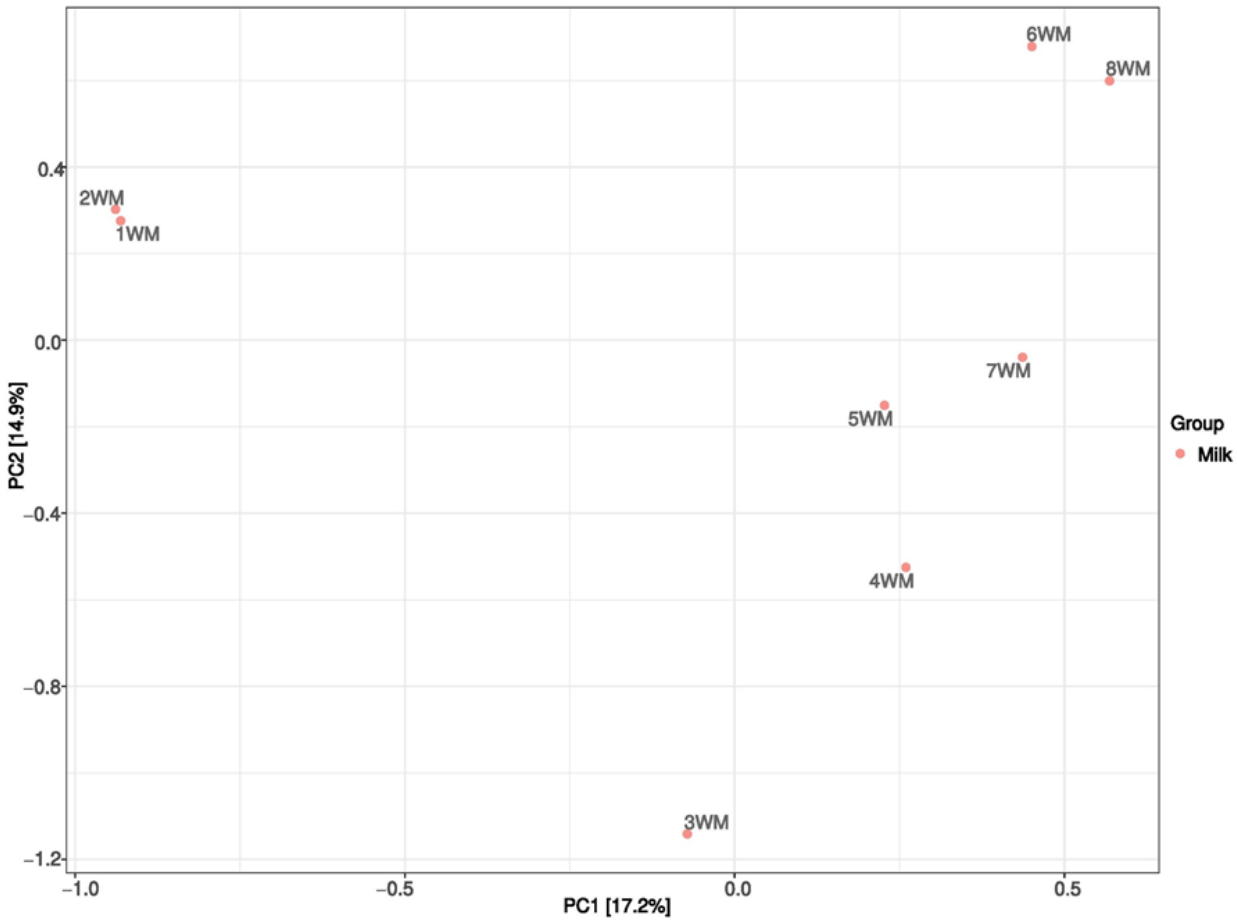

Fig 4.10: β-diversity of milk samples shown with PCA using Hellinger distance metric.

Similar to faecal samples, the core microbial taxa found in bovine milk were catalogued as genus ranks that were consistently present among ≥ 75% of all samples with an overall relative group abundance of ≥ 0.1%. This microbiota consisted of 23 taxa (Table 4.8) accounting for 45.9% of the microbiota obtained in milk. This is quite low when the total number of genus ranking taxa

(374) characterized in milk samples is taken into consideration. Of the 23 core microbial taxa, 7 could be detected in 100% of the milk samples ranging in relative abundance between 1.1 - 7.6%.

Table 4. 8: Core microbiota present in ≥ 75% of milk samples at ≥ 0.1% group abundance including average and range across cow samples

Core microbiota	Number of sequences per genus	Relative abundance (%)	Number of samples positive out of 8 (%)	Average & (range) across milk samples (%)
UCG-005	4948	7.6	8 (100)	4.4 (1.3 – 8. 5)
Bacteroides	2652	4.1	8 (100)	2.4 (0.6 – 4. 7)
Rikenellaceae_RC9_gut_group	2638	4.0	7 (87.5)	2.3 (0.0 – 6.3)
Christensenellaceae_R-7_group	2163	3.3	7 (87.5)	1.8 (0.0 – 2.7)
Prevotellaceae_UCG-004	2049	3.1	7 (87.5)	1.7 (0.0 – 7.4)
Alistipes	1593	2.4	8 (100)	1.5 (0.4 – 4.1)
Akkermansia	1551	2.4	7 (87.5)	1.3 (0.0 – 3.6)
Fusobacterium	1512	2.3	7 (87.5)	1.7 (0.0 – 3.2)
Pseudomonas	1438	2.2	8 (100)	2.6 (0.3 –13.4)
Monoglobus	1362	2.1	8 (100)	1.4 (0.5 – 2.8)
Rhodococcus	956	1.5	7 (87.5)	0.8 (0.0 – 1.9)
NK4A214_group	934	1.4	8 (100)	1.0 (0.3 – 1.9)
Acinetobacter	737	1.1	7 (87.5)	1.0 (0.0 – 4.0)
Phascolarctobacterium	734	1.1	7 (87.5)	1.2 (0.0 – 6.0)
Porphyromonas	719	1.1	8 (100)	0.8 (0.1 – 1.8)
Sphingomonas	656	1.0	6 (75)	1.0 (0.0 – 3.3)
Escherichia/Shigella	627	1.0	7 (87.5)	0.9 (0.0 – 3.9)
dgA-11_gut_group	611	0.9	6 (75)	0.5 (0.0 – 2.2)
Nocardioides	595	0.9	6 (75)	0.6 (0.0 – 2.2)
Treponema	563	0.9	6 (75)	0.5 (0.0 – 1.2)
Iamia	341	0.5	6 (75)	0.3 (0.0 – 0.8)
Luteimonas	318	0.5	6 (75)	0.3 (0.0 – 0.7)
Family_XIII_AD3011_group	317	0.5	6 (75)	0.3 (0.0 – 0.5)

Genus-level taxa forming the core milk microbiota in descending order of abundance of sequences and relative abundance. The number of cows that were positive for each genus,

average and range of the total bacterial population represented by each genus across all cows sampled is also shown in the table.

4.3. BLOOD MICROBIAL COMPOSITION

4.3.1. Sequence analysis

A total yield of 238 434 Illumina reads were obtained from the pooled blood samples (n = 8). The total number of non-chimeric reads obtained including $\bar{x}$ ± SEM were 90 359 (11 295 ± 2581). The number of reads per sample pool ranged from 925 to 26 674 (median 9521; SD 7301).

The overall number of ASVs detected in blood totaled 1747, with 62% of all sequences assigned to genus level taxonomy. Similar to faecal and milk samples, several genera had multiple ASVs associated with them and only one ASV could be resolved to species level. Collapsing of the ASVs yielded a minimum of 4 (5WB) and a maximum of 74 (6WB) genus level taxa that were detected per sample pool.

4.3.2. Alpha and beta diversity analysis

Alpha diversity of blood samples was estimated with three indices *i.e.* Chao1, Shannon and Simpson's estimators. The obtained values are summarized in Table 4.9 showing the minima, maxima, SD and $\bar{x}$ ± SEM for the processed blood samples (n = 8). Similarly to milk, blood samples had a variably rich microbial composition with extremes observed between the highest and lowest richness estimates. High diversity index values were also recorded among blood samples showing the high diversity of the taxa contained within.

Table 4. 9: Summary of alpha diversity estimates within blood samples

Blood samples (n = 8)	**Raw Reads**	**Chao1**	**Shannon**	**Simpson's**
Minimum	8617	18	2.691	0.920
Maximum	18173	1020	6.403	0.998
SD	2671	307	1.052	0.024
$\bar{x}$ ± SEM	14522 ± 944	300 ± 109	4.814 ± 0.372	0.982 ± 0.008

$\bar{x}$ = mean; SD = standard deviation; SEM = standard error of the mean.

The microbial communities hosted in blood samples belonged to three kingdoms including Bacteria (82.5%), Eukaryota (16.9%) and Archaea (0.7%), (Fig 4.11). Eukaryota and Archaea were filtered out for subsequent downstream analyses.

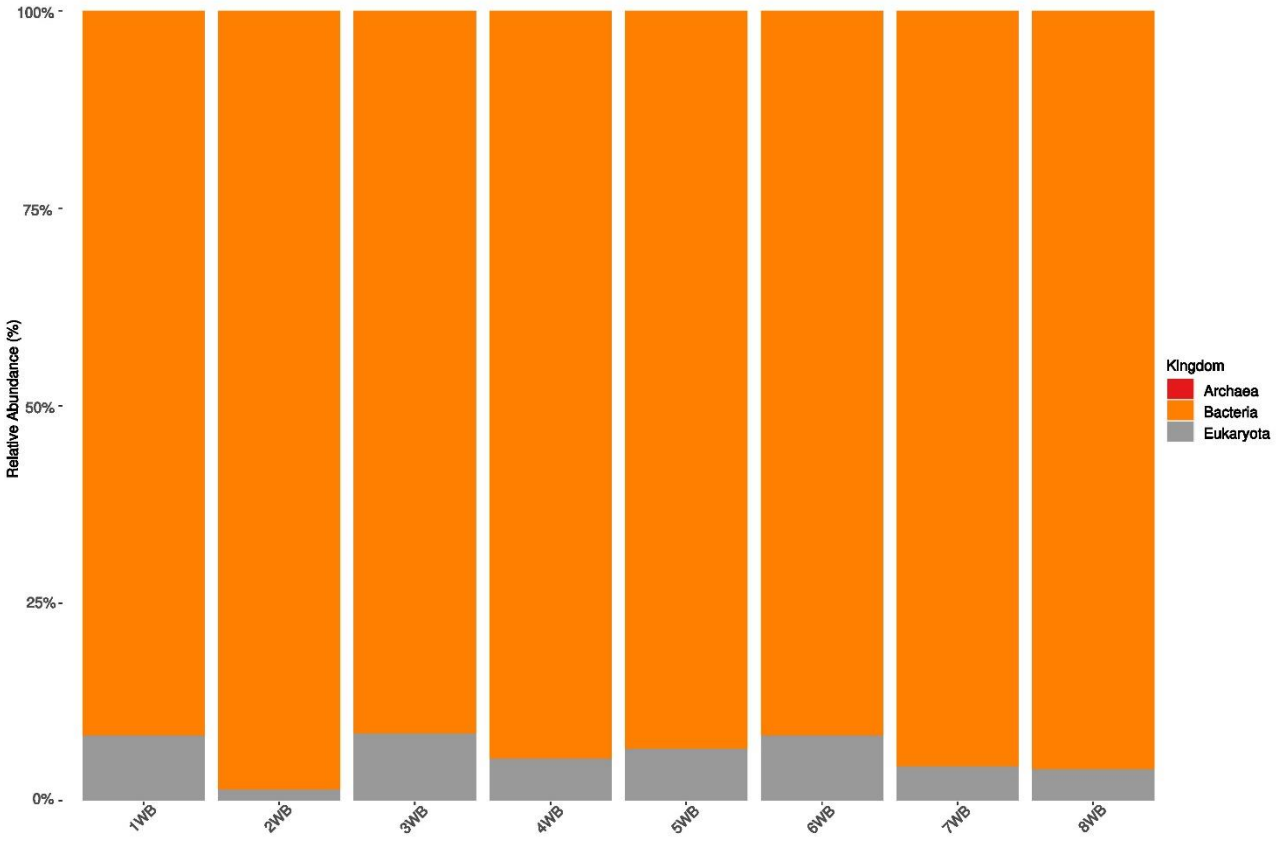

Fig 4.11: Relative abundance chart of the kingdom taxa (legend) detected in blood samples (x-axis).

Eighteen (18) phyla were present in blood samples, with the 5 most abundant being Proteobacteria (66.4%), Firmicutes (20.6%), Bacteroidota (9.4%), Verrucomicrobiota (1.2%) and Actinobacteriota (0.7%), respectively (Fig 4:12). These accounted for 98.3% of the bacterial sequences contained in blood.

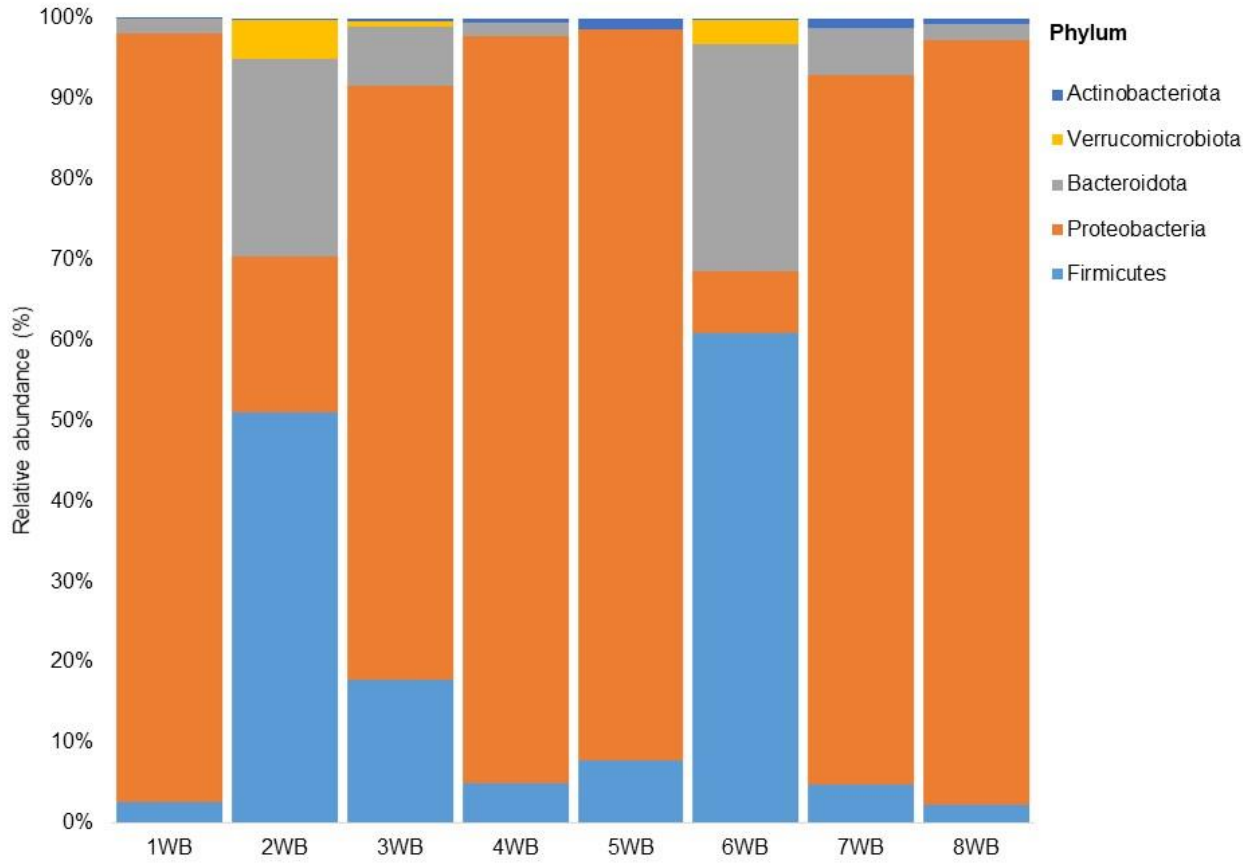

Fig 4.12: Relative abundance chart of the top 5 bacterial phyla (legend) in blood samples (x-axis).

A total of 27 classes were obtained including Alphaproteobacteria (63.9%); Clostridia (17.2%); Bacteroidia (9.4%); Bacilli (3.2%) and Gammaproteobacteria (2.4%) as the top five most abundant.

The number of orders obtained in blood samples were 59, with the five most abundant being Rickettsiales (62.9%); Oscillospirales (11.3%); Bacteroidales (9.4%); Mycoplasmatales (1.7%) and Pseudomonadales (1.6%).

A total of 80 families were detected with Anaplasmataceae (64.5%); UCG-010 (6.6%); Oscillospiraceae (3.3%); Rikenellaceae (3.2%) and Prevotellaceae (2.1%) recorded among the top five most abundant, shown on Fig 4.13.

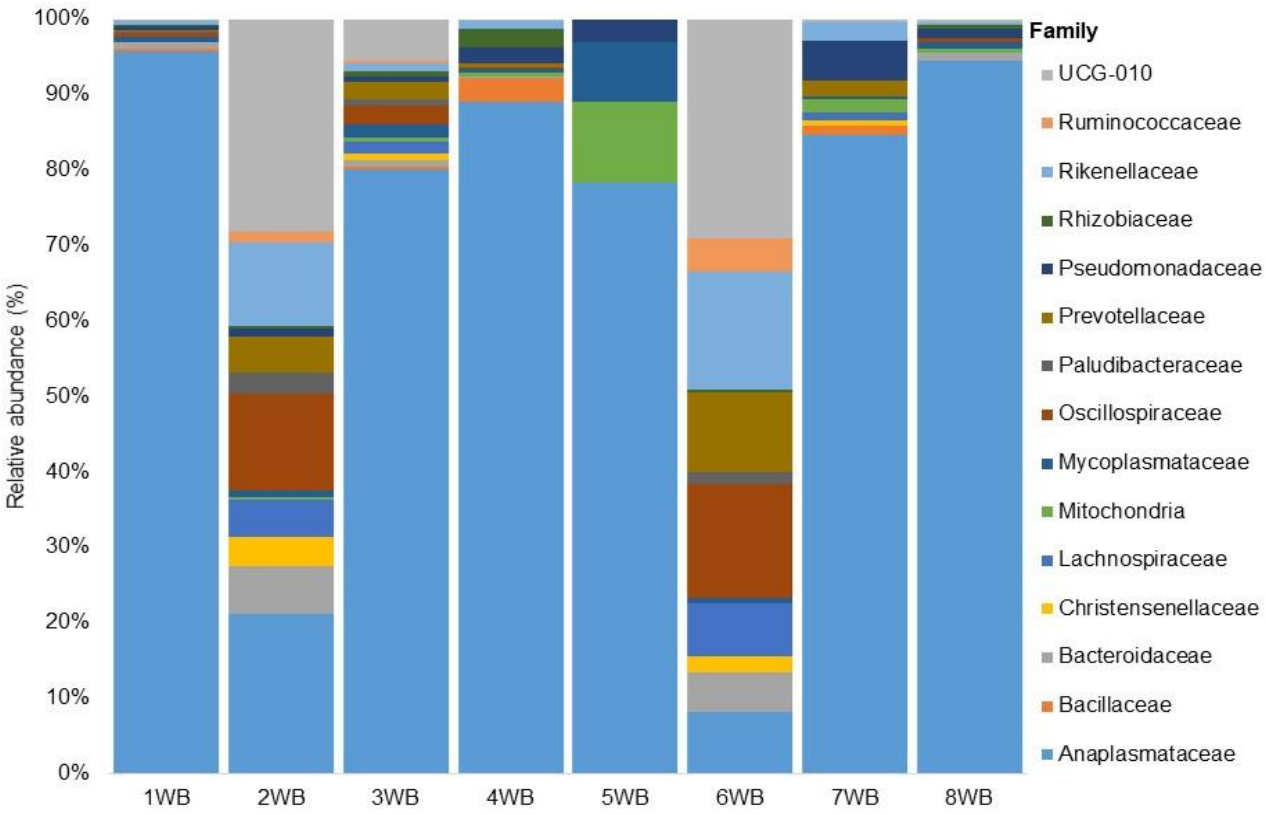

Fig 4.13: Relative abundance chart of the top 15 bacterial families (legend) in blood samples (x-axis).

A total of a hundred and twenty (120) genus ranking taxa could be resolved. The top 15 most abundant genus-level taxa in blood were *Anaplasma* (74.3%); *UCG-005* (2.5%); *Mycoplasma* (2.0%); *Rikenellaceae_RC9_gut_group* (2.0%); *Bacteroides* (1.8%); *Pseudomonas* (1.6%); *Alistipes* (1.6%); *Prevotellaceae_UCG-004* (1.2%); *Christensenellaceae_R-7_group* (0.8%); *Prevotellaceae_UCG-003* (0.6%); *Bacillus* (0.5%); *Monoglobus* (0.5%); *Escherichia/Shigella* (0.4%); *Bartonella* (0.4%) and *Akkermansia* (0.4%). The distribution of these taxa across the blood samples is shown on Fig 4.14 and Table 4.10. The 15 genus ranking taxa accounted for 90.3% of the assigned genera present in the blood samples. The dominant taxa were represented by members of Firmicutes (33%) and Bacteroidota (33%), however Proteobacteria representing one taxon (*Anaplasma*) contributed the most to the overall relative abundance, far surpassing the abundance of members of Firmicutes and Bacteroidota among the blood samples.

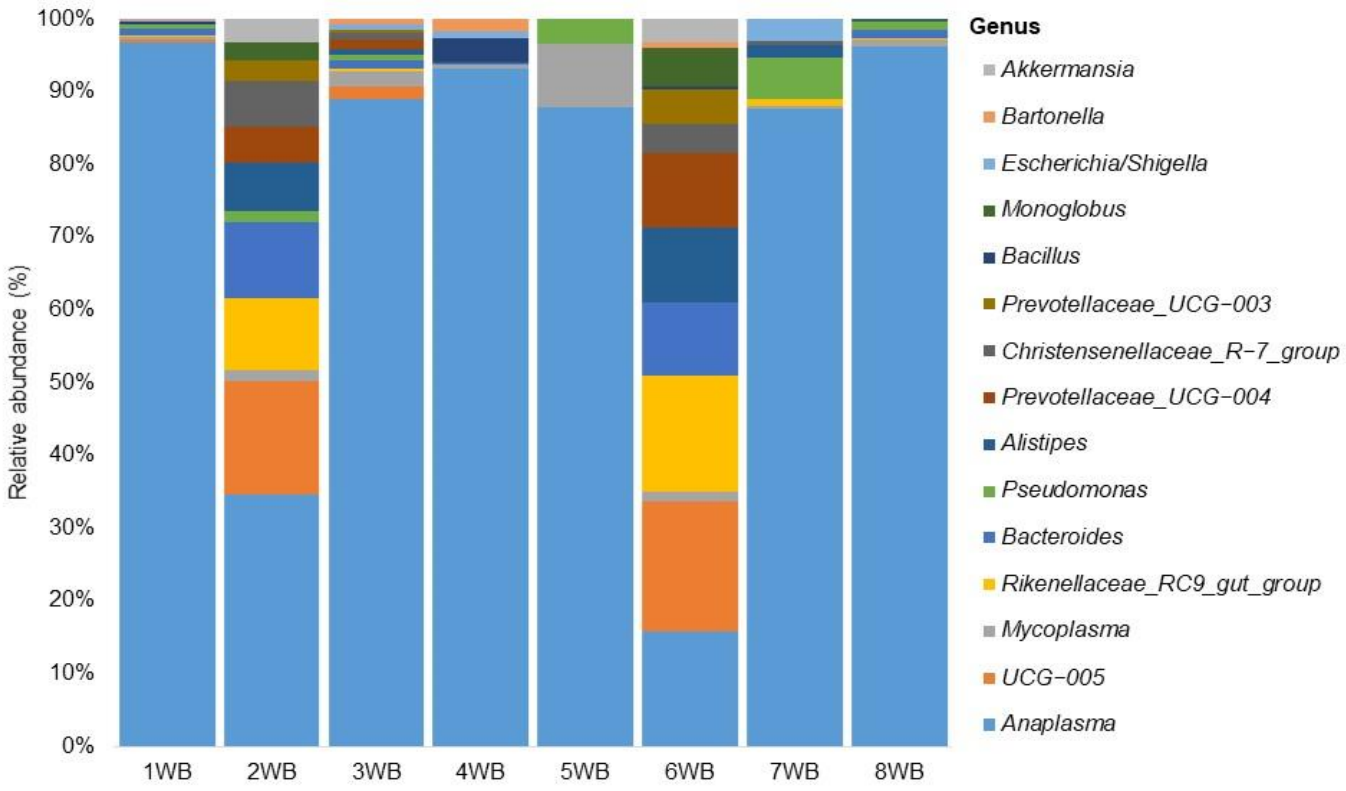

Fig 4.14: Stacked bar plot showing genus-level relative abundance of the top 15 taxa present in blood samples.

Table 4. 10: Distribution of the top 15 most abundant taxa at genus level across blood samples

Genus	Group abundance	1WB	2WB	3WB	4WB	5WB	6WB	7WB	8WB
(P) ***Anaplasma***	69114	15451	**2126**	9545	12292	11813	**961**	5494	11432
(F) ***UCG−005***	2335	55	956	192	0	0	1092	0	0
(F) ***Mycoplasma***	1894	80	92	215	68	1200	81	24	107
(B) ***Rikenellaceae_RC9_gut_group***	1829	29	611	37	0	0	964	53	31
(B) ***Bacteroides***	1682	134	648	119	31	0	616	0	134
(P) ***Pseudomonas***	1505	88	93	93	286	450	6	354	135
(B) ***Alistipes***	1255	0	400	81	28	0	620	109	17
(B) ***Prevotellaceae_UCG−004***	1153	0	313	145	0	0	631	0	0
(F) ***Christensenellaceae_R−7_group***	753	0	378	95	0	0	236	44	0
(B) ***Prevotellaceae_UCG−003***	521	0	177	48	0	0	296	0	0
(F) ***Bacillus***	507	63	0	0	430	0	14	0	0
(F) ***Monoglobus***	494	0	148	0	0	0	323	0	23
(P) ***Escherichia/Shigella***	411	41	0	74	110	0	0	186	0
(P) ***Bartonella***	408	32	0	87	237	0	52	0	0
(V) ***Akkermansia***	400	0	203	0	0	0	197	0	0

Phylum indicated by letter(s) in parentheses before the taxon: B = Bacteroidota, F = Firmicutes, P= Proteobacteria, V= Verrucomicrobiota.

It was observed that the majority of the blood samples (62.5%) had similar community profiles however with great variations in terms of contained taxa. The distribution of the most abundant taxa in these samples also differed quite significantly. Samples 5WB, 2WB and 6WB had the greatest variations that stood out from the rest. Sample 5WB lacked the majority of the top 15 most abundant taxa containing only three genera *i.e*. *Anaplasma, Pseudomonas* and *Mycoplasma.* Overall this sample only contained four microbes assigned at genus level, therefore in addition to the above-mentioned, *Gardnerella* was the fourth. In contrast, samples 2WB and 6WB seemed to host similar communities to each other in more or less proportionate amounts but different from the rest of the samples as shown in red on Table 4.10. Furthermore, the most abundant taxon, *Anaplasma,* registered the lowest counts of abundance in samples 2WB and 6WB when compared to other blood samples (bold text, Table 4.10). Lastly, these two samples contained the greatest number of taxa when compared to the other blood samples.

Anaplasma was present in 100% of all blood samples in disproportionately high abundances compared to other taxa. In addition to *Anaplasma,* genera of veterinary significance that listed among the top 15 most abundant taxa include *Mycoplasma, Bacteroides*, *Pseudomonas, Bacillus, Escherichia/Shigella* and *Bartonella.* Other noteworthy taxa of veterinary significance detected in blood samples in lower abundances were *Fusobacterium* (0.3%); *Ehrlichia* (0.2%); *Streptococcus* (0.1%); *Peptostreptococcus* (0.1%); *Prevotella* (0.1%); *Rhodococcus* (0.1%), *Klebsiella* (<0.1%) and *Staphylococcus* at (<0.1%).

Mycoplasma and *Pseudomonas* were detected in 100% of the blood samples ranking third and sixth. *Bacteroides* was detected in 75% of the samples, except for sample 5WB and 7WB; while *Bartonella* and *Escherichia/Shigella* were both detected in four (50%) samples each, *i.e.* 1WB, 3WB, 4WB and 6WB – *Bartonella* and samples 1WB, 3WB, 4WB and 7WB – *Escherichia/Shigella*. *Bacillus* was detected in three (37.5%) samples (1WB, 4WB and 6WB); while *Prevotella* was only detected in two (25%) samples (6WB& 7WB).*Staphylococcus* was detected in one (12.5%) sample (8WB); likewise, *Ehrlichia* (7WB), *Rhodococcus* (3WB), *Peptostreptococcus* (1WB) and *Klebsiella* (1WB). There was no species-level resolution of taxa among the blood samples.

Multivariate ordination analysis using PCA was used to assess the β-diversity of the microbial populations found in each sample (Fig 4.15). The majority of the samples clustered together except for sample 2WB, 5WB and 6WB. From the observations above, this deviation was expected since this trio varied considerably in terms of their microbial composition. The visual

observation based on principal component ordination supports the findings plotted on the stacked bar plots representing genus ranking taxa (Fig 4.14). The first principal component accounted for 30.9% of the variation. The expectation was also that sample 2WB and 6WB would cluster closely together due to the similarities in their microbial constituents with regards to the top 15 most abundant genera. Further analysis into their individual constituent taxa showed that the two sample pools had 37 uniquely shared taxa between them and 53 taxa being variably present in either 2WB or 6WB; 6WB bearing more taxa than 2WB including highly abundant *Bacillus* and *Bartonella*. In addition to other determinants (inevident in this analysis) it can thus be speculated that this could be the reason for the divergence between the two.

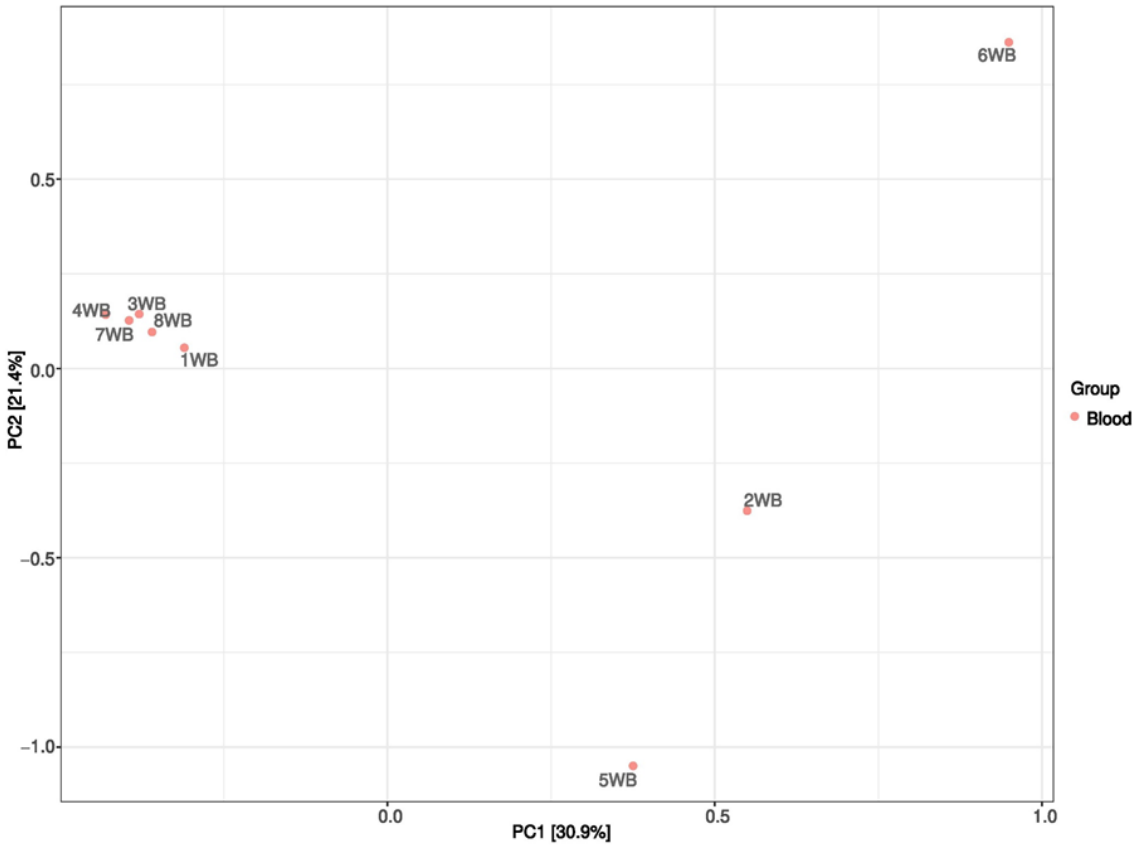

Fig 4.15: β-diversity of blood samples shown with PCA on Hellinger distance metric.

Similarly to faecal and milk samples, the core microbial taxa found in bovine blood were catalogued as genus ranks that were consistently present among ≥ 75% of all samples with an overall relative group abundance of ≥ 0.1%. This microbiota consisted of 6 taxa accounting for 83% of the taxa obtained in blood. These are listed in descending order of abundance on Table 4.11. Only three genus ranking taxa were present across 100% of the blood samples *i.e*. *Anaplasma*, *Mycoplasma* and *Pseudomonas*.

Table 4. 11: Core microbiota present in ≥ 75% of blood samples at ≥ 0.1% group abundance including average and range across cow samples

Core microbiota	Number of sequences per genus	Relative abundance (%)	Number of samples positive out of 8 (%)	Average & (range) across blood samples (%)
Anaplasma	69114	74.3	8 (100)	58.9 (5.3 – 92.3)
Mycoplasma	1894	2.0	8 (100)	2.1 (0.3 – 7.8)
Rikenellaceae_RC9_gut_group	1829	2.0	7 (87.5)	1.6 (0.0 – 5.3)
Bacteroides	1682	1.8	6 (75)	1.5 (0.0 – 4.2)
Pseudomonas	1505	1.6	8 (100)	1.4 (0.0 – 4.1)
Alistipes	1255	1.3	6 (75)	1.2 (0.03 – 3.4)

Genus-level taxa forming the core blood microbiota in descending order of abundance of sequences and relative abundance. The number of cows that were positive for each genus, average and range of the total bacterial population represented by each genus across all cows sampled is also shown in the table.

4.4. PLACENTAL MICROBIAL COMPOSITION

4.4.1. Sequence analysis

Coincidentally, placental tissue samples were obtained during sample collection. Three different portions of the tissue were snipped paying particular attention to the cotyledons. These were then processed and homogenized into one placental sample for NGS analysis. The number of non-chimeric reads retained in the placental tissue was 261 740. Overall 271 ASVs were obtained and 270 (99.6%) were assigned to genus level. The ASVs were collapsed to a total of 16 genera. Two ASVs were resolved to species level in the placental tissue.

4.4.2. Abundance & diversity

The abundance according to Chao 1 index estimator for the placental tissue was 368.11. The tissue microbiota was constituted of 100% Bacteria. There were five phyla detected, listed here in descending order of abundance *i.e.* Fusobacteriota (43.2%), Bacteroidota (34.9%), Firmicutes (15.9%), Actinobacteriota (3.6%) and Proteobacteria (2.4%).

A total of 8 classes which included Fusobacteriia (43.2%), Bacteroidia (34.9%), Clostridia (14.4%), Actinobacteria (3.5%) and Bacilli (1.5%) as the top five most abundant were resolved.

The taxa could be clustered into ten orders which included Fusobacteriales (43.2%); Bacteroidales (34.9%), Peptostreptococcales-Tissierellales (12.5%); Actinomycetales (3.4%) and Lactobacillales (1.5%) as the top five most abundant.

Out of the 14 families obtained in placental tissue, the top 5 most abundant included Fusobacteriaceae (43.2%), Bacteroidaceae (18.3%), Porphyromonadaceae (16.6%), Anaerovoracaceae (5.0%) and Peptostreptococcaceae (3.8%) in decreasing order of abundance.

A total of 16 taxa obtained in the placental tissue were resolved to genus level at ≥ 0.1% relative abundance. The ten most abundant genera of bacteria in descending order were *Fusobacterium* (43.1%), *Bacteroides* (18.3%), *Porphyromonas* (16.6%), *S5-A14a* (5.0%), *Peptostreptococcus* (3.8%), *Helcococcus* (3.5%), *Trueperella* (2.0%), *Streptococcus* (1.5%), *Actinomyces* (1.5%) and

Mannhemia (1.4%). These and the remainder of the genera detected in the placental tissue are listed on Table 4.12.

Table 4. 12: Genus-level microbiota detected in placental tissue at ≥ 0.1% relative abundance.

Genus	Group Abundance	Relative abundance (%)
Fusobacterium	10544	43.1
Bacteroides	4471	18.3
Porphyromonas	4061	16.6
S5-A14a	1220	5.0
Peptostreptococcus	930	3.8
Helcococcus	852	3.5
Trueperella	487	2.0
Streptococcus	367	1.5
Actinomyces	357	1.5
Mannheimia	349	1.4
Brucella	236	1.0
Anaerococcus	35	0.1
Peptoniphilus	28	0.1
Cetobacterium	20	0.1
Rhodococcus	19	0.1

The microbial community structure of the placental tissue is depicted on Fig 4.16. The heat map shows bacterial taxa contained within placental tissue in descending order of abundance, including an unclassified taxon in the phylum Actinobacteriota.

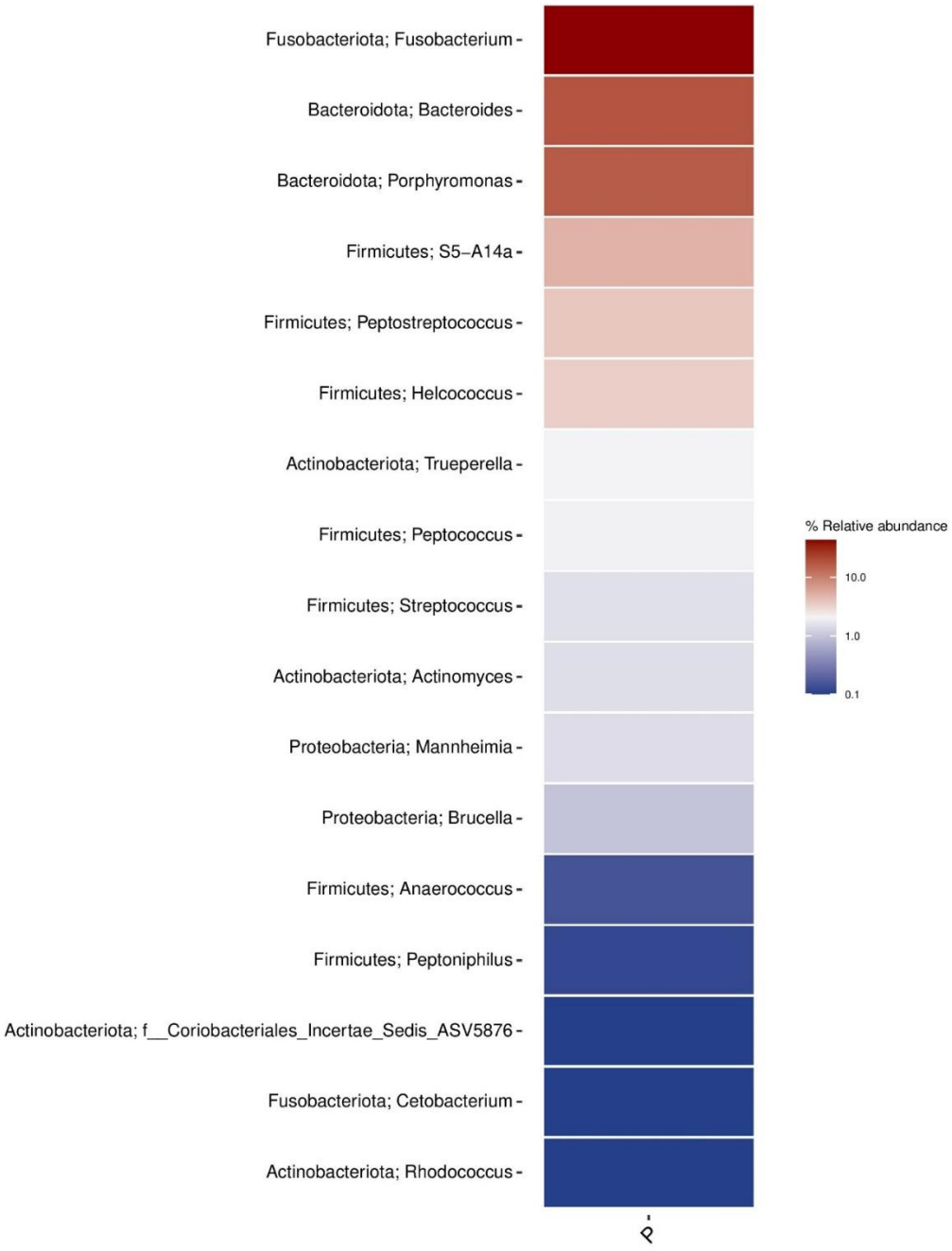

Fig 4.16: Genus-level taxa obtained in placental tissue, heat map plot depicts the relative abundance in percentage of each bacterial taxon in the sample. The relative values for the microbial taxa are depicted by colour intensity in the legend indicated on the *right* of the figure. The sample pool (P) shown along the *x-axis* and the contained taxa along the *y-axis*, respectively.

The sample pool was largely dominated by important microbial taxa primarily *Fusobacterium*; secondarily *Bacteroides* and *Porphyromonas.* A notably important genus, *Brucella,* with a relative abundance of 1.0% was present among these. Furthermore, other genera associated with bovine diseases and/ or conditions including *Helcococcus*, *Trueperella, Streptococcus, Peptoniphilus* and *Rhodococcus* were also detected. The ASVs that were resolved to species level taxonomy

in placental tissue were *F. necrophorum* (ASV149) and *Peptostreptococcus anaerobius* (ASV2906).

4.5. COMPARATIVE MICROBIAL DIVERSTY AND COMPOSITION

4.5.1. Comparative microbial diversity and composition of corresponding faeces, milk and blood samples

4.5.1.1. Sampling results and sequence analysis

A total of 24 pooled samples were processed for comparative downstream analysis, constituted by faeces (n = 8), milk (n = 8) and blood (n = 8). The mean read length of sequences was 430 bp (min = 251 bp; max = 468 bp). A total of 1 606 440 Illumina reads were generated from 911 912 446 base yield resulting from sequencing of the V3 - V4 hypervariable region of the 16S rRNA gene amplicons of pooled faecal, milk and blood samples. A total of 602 011 (min = 925; mean = 25 084; max = 58 202; SD = 16 836) non-chimeric reads were retained.

An average of 44 424 reads were obtained in faeces, 19 532 in milk and 11 295 in blood. The least number of reads (925) were obtained in blood (5WB), while the highest (58 202) were in faeces (7WF).

The reads in all three sample groups were assembled into 8426 distinct ASVs at kingdom Bacteria level through DADA2 pipeline. At least 98.8% of the sequences could be assigned to a known phylum, with the proportion of assignments decreasing at lower taxonomic levels (Fig 4.17). Several taxa were assigned multiple ASVs as the pipeline can resolve differences in sequence variants at as low as a single nucleotide. The least count of ASVs was observed among blood samples (18), while the highest count was among faecal samples (1977); both samples were from the same group of animals (5WB & 5WF). The greatest sum total of unique ASVs were recorded in faeces (12 632), followed by milk (4 764) and then blood (2 529). Overall the ASVs were collapsed into a minimum of 5 (5WB) and maximum of 331 (6WM) ASVs representing unique microbial taxa per sample pool. A minimum of 4 and maximum of 207 of these microbial taxa contained per sample pool were assigned at genus level taxonomy (see Table 4.2).

A total of 67.2% of the taxa could be resolved to genus level, while 28.5% could not. The remaining 4.3% at this level was unaccounted for. Only three ASVs were resolved at species level among the three niches *i.e*. *F. necrophorum, L. composti* and *R. sedimentorum*, accounting for a small fraction (0.027%) of species present in bovine faeces, milk and blood samples. Although the genus *Fusobacterium* was detected in all niches, the species *F. necrophorum* was only

detected among milk samples (50% of samples). The genus *Luteimonas* was only detected across milk samples (75% of samples), but *L. composti* was resolved from one milk sample (12.5%). The species *R. sedimentorum* was detected only among faecal samples (100% of samples) although the genus did also occur in blood (12.5% of samples) and in milk (37.5% of samples).

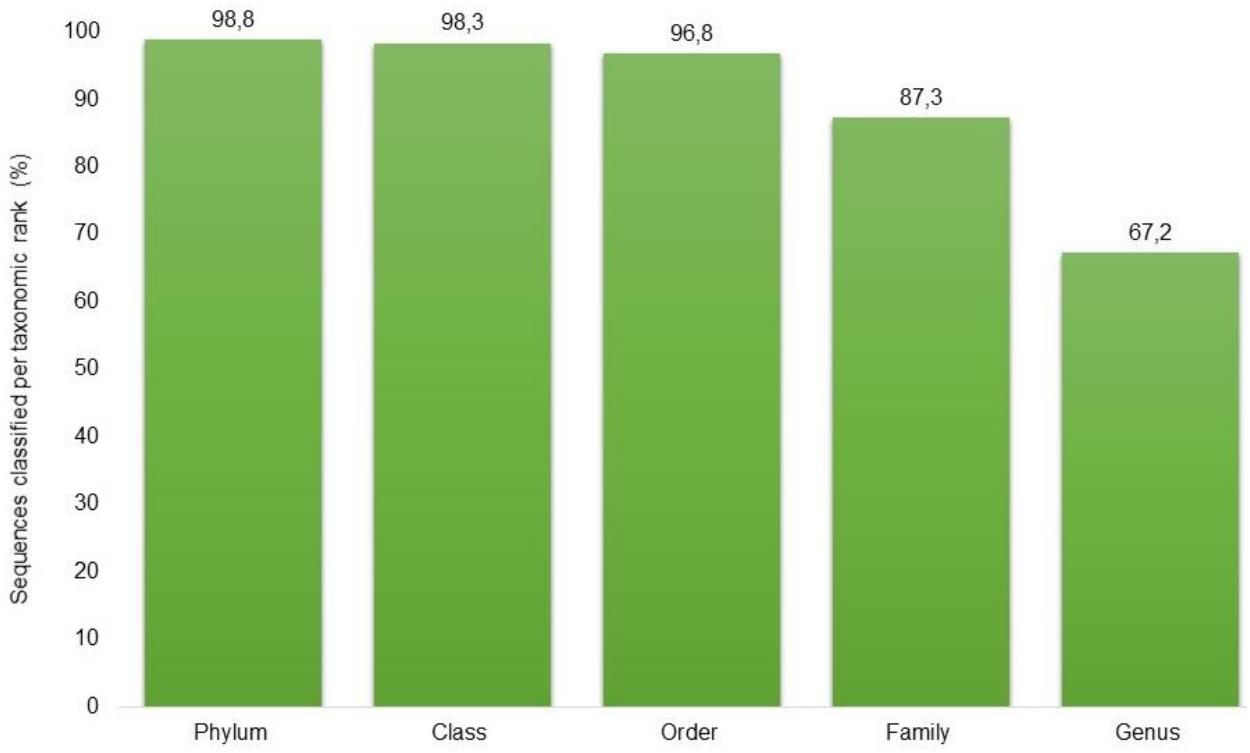

Fig 4.17: The proportion of amplicon sequence variants (ASVs) assigned at a given taxonomic rank using the SILVA database v138.

In order to determine the similarities and / or differences between the the structure of the microbiota contained within faeces, milk and blood samples collected from the same group of animals, a comparative analysis of the constituent microbes was conducted. This was achieved through determination of the alpha and beta diversity using various bioinformatics tools.

4.5.1.2. Alpha and beta diversity analysis

When analyzing differences between the three sample groups, the alpha diversity box-plots intuitively reflected the minima, median, degree of dispersion, maxima, and outliers of microbial diversity within groups (Fig 4.18 A-C). The alpha diversity of the three sample groups was estimated by ASV richness and diversity indices *i.e.* Chao1, Shannon & Simpson's. The microbial communities from the faecal samples had significantly higher alpha diversity than milk and blood samples as determined through the three index estimators respectively (Kruskal-Wallis: p = 0.00081, p = 0.001, & p = 0.0031; significant at $p < 0.05$).

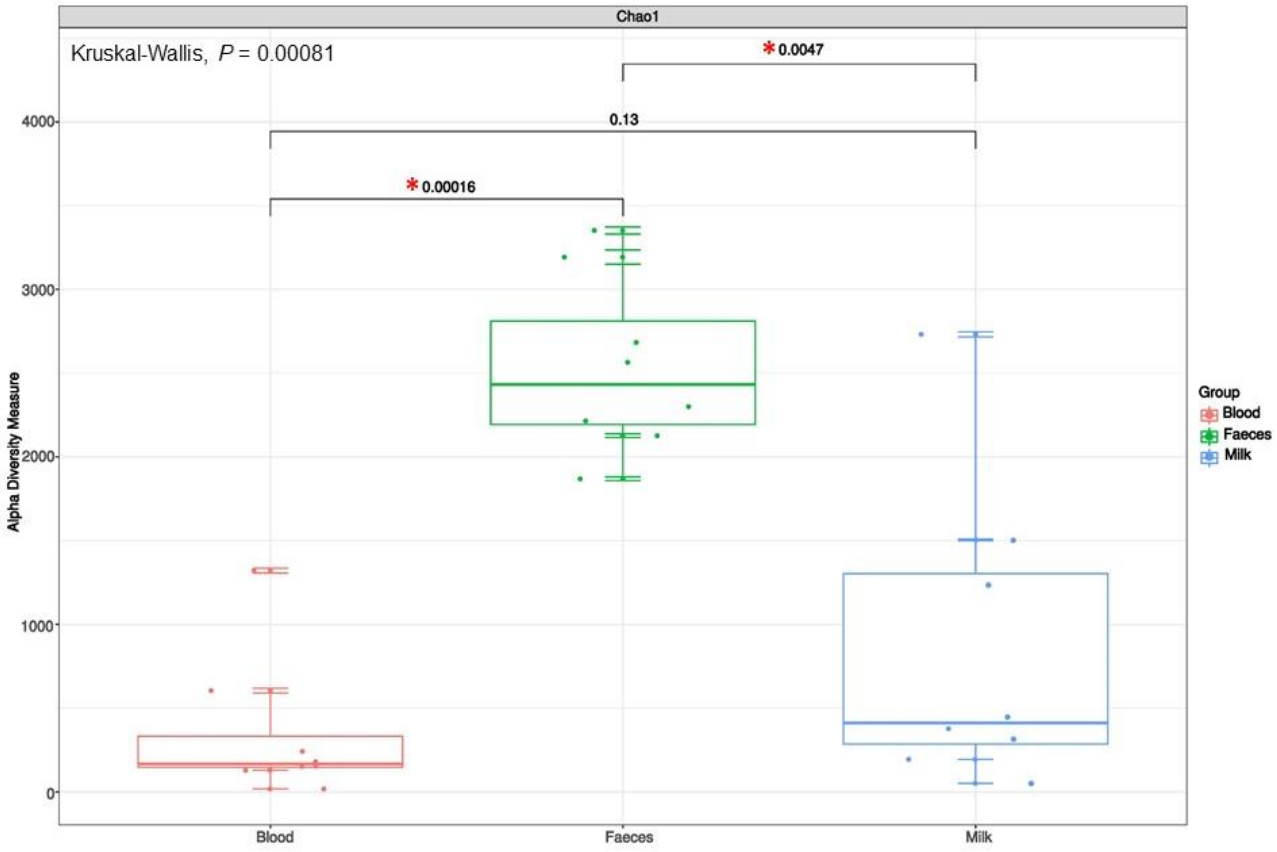

Fig 4.18 A: Alpha diversity box-plots showing Chao1 species richness estimates per sample group *i.e.* blood (red), faeces (green) and milk (blue). *Significant at $p < 0.05$.

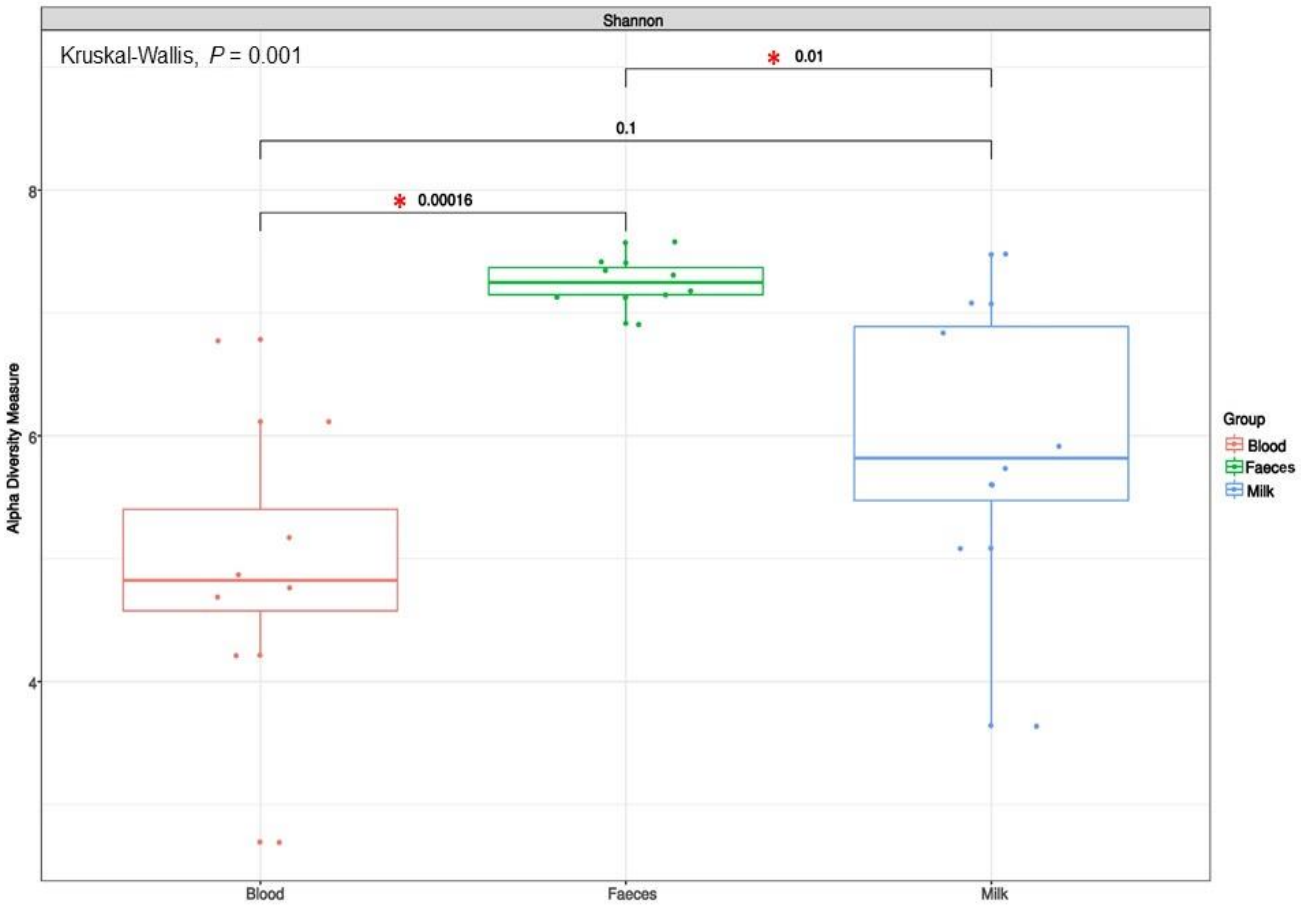

Fig 4.18 B: Alpha diversity box-plots showing Shannon species diversity estimates per sample group *i.e.* blood (red), faeces (green) and milk (blue). *Significant at $p < 0.05$.

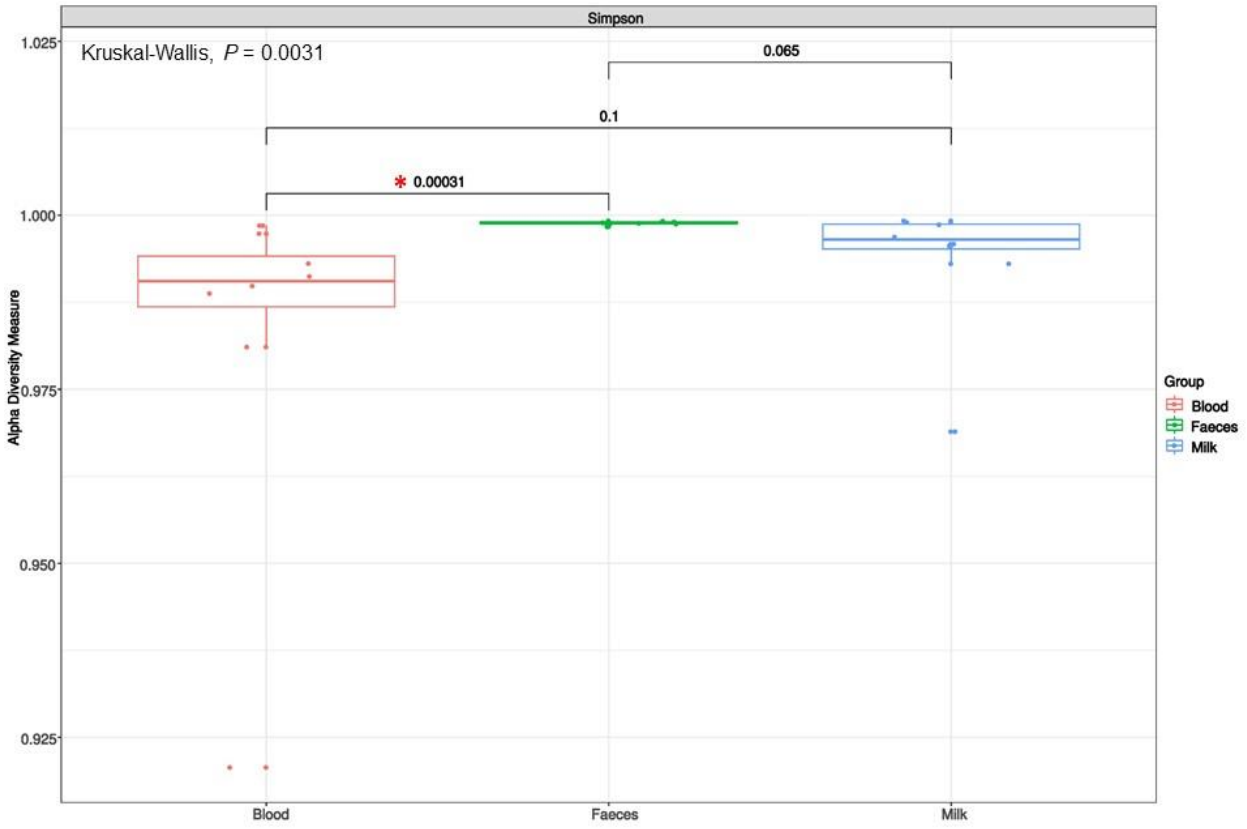

Fig.4.18 C: Alpha diversity box-plots showing Simpson's species diversity (evenness) estimates per sample group *i.e.* blood (red), faeces (green) and milk (blue). *Significant at $p < 0.05$.

The estimated species richness measured through Chao1 index, varied from 18 to 1710.49 with mean value of 784.93. Comparison of the Shannon indices across all samples showed microbial diversity between the sample groups with the following values: min = 2.69; mean = 5.67 and max = 6.73. The Simpson's diversity index ranged from 0.92 to 1. The values indicate high microbial diversity across the different sample groups.

Overall, faecal samples displayed higher alpha diversity values than both milk and blood, as determined through Chao1, Shannon and Simpson indices. Furthermore, milk microbial communities showed more richness and diversity than those of blood; with blood microbial communities being the least diverse. The α-diversity varied significantly between blood and faeces based on the three index estimators ($P < 0.05$); between faeces and milk groups via Chao1 & Shannon indices ($P < 0.05$) but did not vary significantly between blood and milk via the three index estimators.

Taxonomy bar charts were created to note differences and similarities among taxonomic ranks between the sample groups. All in all, three kingdoms were identified *i.e*. Archaea (in blood, faeces and milk), Bacteria (in blood, faeces and milk) and Eukaryota (in blood and milk). ASVs assigned as Archaea and Eukaryota were subsequently filtered out for further downstream analysis

An overall total of 30 bacterial phyla were present across the three groups of samples, the highest number (30) was detected in milk samples (Table 4.13). Despite the high alpha diversity, faecal samples recorded the least number of phyla (14), surpassed by blood samples which contained 18 phyla yet having the least alpha diversity of the three main sample groups.

There was an increase in the number of taxa with lower ranking. Overall 74 classes were resolved across the three sample groups, the majority (74) found in milk samples and the least in faeces (21). A total of 156 orders could be resolved, while 243 families and 408 taxa representing genera were resolved. The highest number of taxa in the afore-mentioned rankings were identified in milk (154 orders, 236 families and 374 genera), whereas the lowest count of taxa were observed in faeces (38 orders, 55 families and 98 genera). Table 4.13 shows the taxonomic ranking per sample group and overall number of ranks detected in the entire analysis when taking into consideration the three main sample groups.

Table 4. 13: Microbial taxonomic ranking and overall number of ranks per sample group

Group	Kingdom*	Phylum	Class	Order	Family	Genus
Faeces	2	14	21	38	55	98
Milk	3	30	74	154	236	374
Blood	3	18	27	59	80	120
Overall number of ranks	3	30	74	156	243	408

*Only kingdom level classification includes members of Archaea and Eukaryota, other downstream analyses were performed strictly on Bacteria.

Overall, the five most abundant phyla that were common across all three groups of samples were Firmicutes (42.5%), Proteobacteria (25.6%), Bacteroidota (18.8%), Verrucomicrobiota (3.0%) and Actinobacteriota (3.0%) in varying group abundances (Fig 4.19 A). These accounted for 97.6%, 85.3% and 98.3% of the bacterial communities in faeces, milk and blood, respectively. A total of 15 of the 30 phyla (50%) were shared between faeces, milk and blood.

The faecal group was dominated by Firmicutes and Bacteroidota (at 64% and 25.9% relative abundances, respectively); likewise, milk samples (at 39.4% and 20.4% relative abundances, respectively). The blood was predominated by Proteobacteria (66.4%) and Firmicutes (20.6%). Verrucomicrobiota and Actinobacteriota were also abundant in milk (3.2% and 7.3%, respectively) and faeces (4.4% and 1.7%, respectively) but less abundant in blood (1.2% and 0.7%, respectively). Milk also contained members of Proteobacteria (15.0%), with a negligible amount in faeces (0.2%). This distribution of phyla between samples is shown on (Fig 4.19 B)

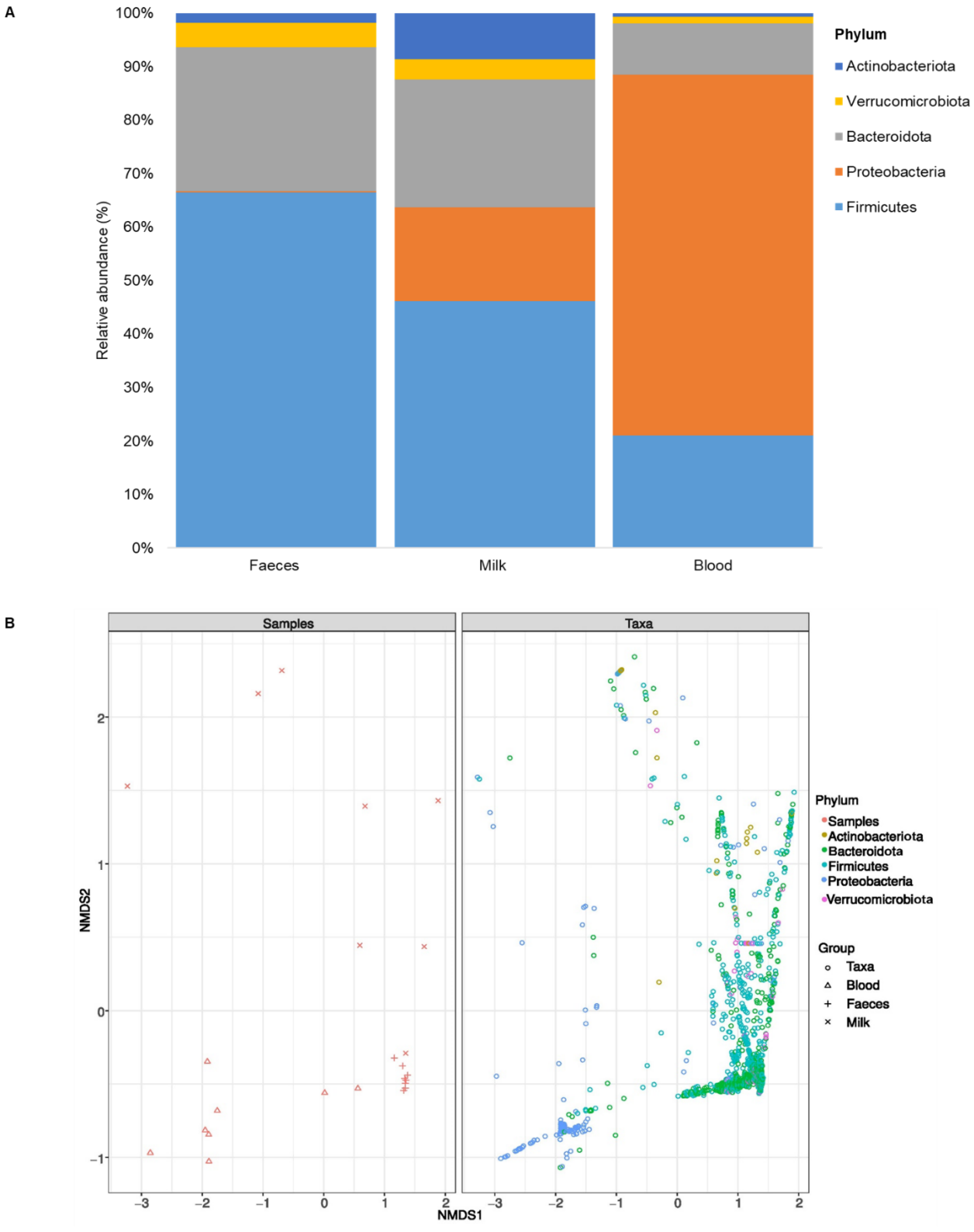

Fig 4.19 A & B: Distribution of the five most abundant phyla. A: Stacked bar-plot with proportions of bacterial phyla detected from the three sample groups, relative abundance graphed along the y-axis and sample type along the x-axis. B: NMDS split bi-plot of samples & ASVs at phylum level (legend).

The top five most abundant classes overall in descending order of abundance were Clostridia (38.7%); Alphaproteobacteria (22.2%); Bacteroidia (18.7%); Gammaproteobacteria (3.3%) and Bacilli (3.3%) accounting for 86.2% of the overall relative abundance of classes.

Among the orders, Rickettsiales (20.1%); Bacteroidales (18.2%); Oscillospirales (16.7%); Peptostreptococcales-Tissierellales (10.9%) and Lachnospirales (2.6%) listed as the top five most abundant accounting for 68.5% of the overall relative abundance.

At the family level, 243 families (87.3% relative abundance) were detected and the abundance of unclassified bacteria in the samples was 11.3%. The distribution proportion was greater than 0.5% for 30 families. Among these the 15 most abundant families overall in descending order of abundance were as follows: Anaplasmataceae (19.4%); Peptostreptococcaceae (9.2%); Oscillospiraceae (6.3%); UCG-010 (5.9%), Rikenellaceae (5.4%); Prevotellaceae (5.0%); Lachnospiraceae (2.5%); Bacteroidaceae (2.3%), Akkermansiaceae (2.2%); Monoglobaceae (2.1%); Ruminococcaceae (2.0%); Christensenellaceae (1.9%); Anaerovoracaceae (0.9%); Pseudomonadaceae (0.8%) and Erysipelotrichaceae (0.8%), Fig 4.20.

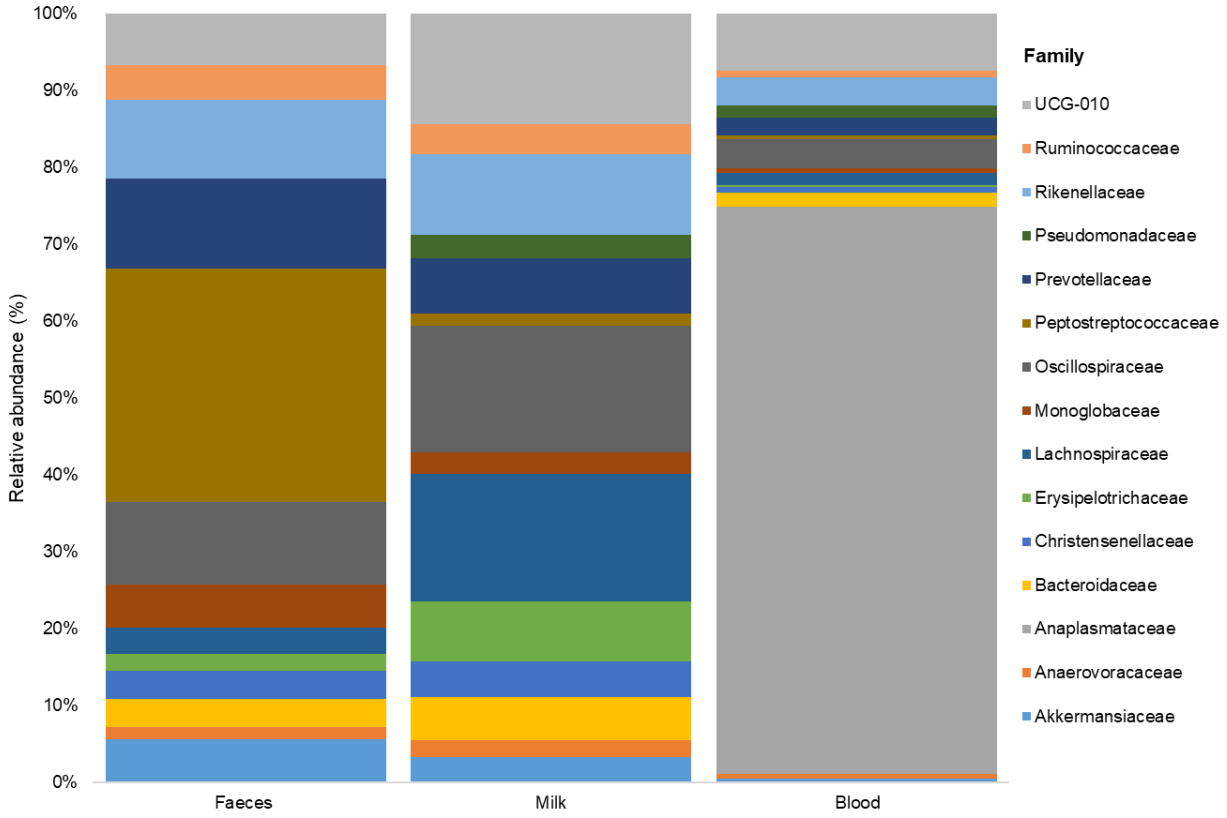

Fig 4.20: Stacked bar-plot with proportions of bacterial families detected from the three sample groups, relative abundance graphed along the y-axis and sample type along the x-axis.

The distribution pattern of the microbial families (Fig 4.20) contained in faeces and milk looks very similar while that of blood is quite unique with Anaplasmatacae being the most dominant family among the blood samples but absent in faeces and milk.

At the genus level, the percentage of unclassified bacteria was 28.5%. Overall, 408 genus-level taxa were identified accounting for 67.2% of the microbial taxa (Annexure A). The distribution proportion was greater than 0.5% in 21 (5.1%) genus-level taxa. The top 15 abundant genus-level taxa with their respective rankings (1 = most abundant; 15 = least abundant) and their distribution across the three niches are shown on Table 4.14. These taxa contributed 51.6 % (83.2%, 34.7% and 87.9% in faeces, milk and blood, respectively) to the overall abundance of the taxa assigned at genus level and it was dominated by members of the phylum Bacteroidota (40%), followed by Firmicutes (33%).

Table 4. 14: Top 15 abundant taxa with their respective overall rankings and distribution across the three main sample groups.

Rank	Top 15 overall genus-level taxa	Overall relative abundance of taxa (%)	Faecal group relative abundance (%)	Milk group relative abundance (%)	Blood group relative abundance (%)	Samples positive for taxa
1	(P) *Anaplasma*	19.4	**0**	**0**	74.3	1-8WB
2	(F) *Romboutsia*	7.4	26.6	0.6	0.3	1-8WF; 2,5-8WM; 3&6WB
3	(F) *UCG−005*	4.4	8.9	7.6	2.5	1-8WF; 1-8WM; 1-4,6&7WB
4	(B) *Prevotellaceae_UCG−004*	3.1	8.3	3.1	1.2	1-8WF; 1-8WM; 2-4,6&7WB
5	(B) *Rikenellaceae_RC9_gut_group*	2.7	5.4	4.0	2.0	1-8WF; 1-3,4-8WM; 1-4,6-8WB
6	(B) *Bacteroides*	2.3	4.0	4.1	1.8	1-8WF; 1-8WM; 1-4,6&8WB
7	(V) *Akkermansia*	2.2	6.0	2.4	0.4	1-8WF; 1-8WM; 2&6WB
8	(F) *Monoglobus*	2.1	6.0	2.1	0.5	1-8WF; 1-8WM; 2,6&8 WB
9	(F) *Christensenellaceae_R−7_group*	1.9	3.9	3.3	0.8	1-8WF; 1-2,4-8WM; 2,3,6&7WB
10	(B) *Alistipes*	1.7	3.4	2.4	1.3	1-8WF; 1-8WM; 2-4,6-8WB
11	(B) *Prevotellaceae_UCG-003*	1.1	2.8	1.2	0.6	1-8WF; 4-8WM; 2,3,6WB
12	(F) *Paeniclostridium*	1.0	3.6	0.3	0.1	1-8WF; 6&7WM; 6WB
13	(P) *Pseudomonas*	0.8	0.03	2.2	1.6	2WF; 1-8WM; 1-8WB
14	(B) *dgA-11_gut_group*	0.8	2.0	0.9	0.4	1-8WF; 1-2,5-8WM; 1-2&6WB
15	(Pa) *Candidatus_Saccharimonas*	0.7	2.3	0.5	0.1	1-8WF; 4-6&8WM; 2-3,6WB

Phylum indicated by letter(s) in parentheses before the taxon: B = Bacteroidota, F = Firmicutes, P = Proteobacteria, Pa = Patescibacteria, V = Verrucomicrobiota.

The most prevalent genus-level taxa detected in the faeces included, as expected, members of the Firmicutes (*e.g. Romboutsia* at 26.6%, *UCG−005* at 8.9% and *Christensenellaceae_R-7_group* at 3.9%) and Bacteroidota (*e.g. Prevotellaceae_UCG−004* at 8.3% and *Rikenellaceae_RC9_gut_group* at 5.4%).

Similarly, bacterial sequences in milk were predominantly from Firmicutes (*e.g. UCG-005* at 7.6% and *Christensenellaceae_R-7_group* at 3.3%) and Bacteroidota (*e.g. Rikenellaceae_RC9_gut_group* and *Bacteroides* at 4.1% and 4.0%, respectively).

On the other hand, blood was predominated by Proteobacteria representing one taxon (*Anaplasma,* 74.3%) in high abundance, however members of Firmicutes (*UCG-005*, 2.5% and *Mycoplasma*, 2.0%) and Bacteroidota (*e.g. Bacteroides*, 1.8%) were also detected and more numerous although in lower abundances.

It could be seen from Table 4.14 that the distribution of the most abundant taxa was quite variable across the different sample groups. In some instances, one group contributed more to the overall abundance than others *e.g.* in the case of *Anaplasma* (74.3%), the blood group was the sole contributor. While in the case of *Romboutsia*, the faecal group contributed the most (26.6%) to the overall abundance compared to milk and blood groups which in turn contributed 0.6% and 0.3%, respectively.

Based on overall observation, the blood group contained all 15 of the most abundant taxa. However, only *Anaplasma* was contained within 100% of the sample pools and the other 14 genus-level taxa were variably present among blood sample pools.

Similarly, this varied distribution pattern was also observed among milk samples. The taxa that were detected in 100% of the milk samples were *UCG-005, Prevotellaceae_UCG-004, Bacteroides, Akkermansia, Monoglobus, Christensenellaceae_R-7_group* and *Alistipes*.

In contrast 100% of the faecal sample pools contained 14 of the 15 most abundant taxa, with the exception of *Pseudomonas* which was only detected in sample 2WF.

Monoglobus under the phylum Firmicutes was unexpected across all sample groups and it was detected at an overall relative abundance of 2.1%, being highest in abundance among the faecal group (6.0%), followed by milk (2.1%) and then the blood group (0.5%).

Paeniclostridium was detected in 100% of faecal samples at 3.6% relative abundance. It was however detected in one sample in the blood group (*i.e*. 6WB at 0.1% relative abundance) and in

two samples in the milk (*i.e*. 6WM and 7WM at 0.3% relative abundance). The genus contributed 1.0% to the overall relative abundance.

Interestingly the genus *Akkermansia* was also among the most abundant taxa in the group comparisons, ranking 7th overall with a relative abundance of 2.2% across the three niches.

The pathogen *Fusobacterium* was also detected among the three sample groups at an overall relative abundance of 0.5%. It was however detected in only one faecal sample *i.e*. 1WF. Worth noting is that this particular sample was obtained from the same animal which had a retained placenta from which *Fusobacterium* was detected at 44.5% relative abundance. The milk and blood of this animal also contained this pathogen.

When investigating the distribution patterns of microbial taxa across the three niches, it could be seen that bacterial profiles from milk more closely resembled those from faeces in terms of contained taxa and proportions of the 15 most abundant genus ranking taxa (Fig 4.21A). Although blood samples also possessed similar taxa, they were in lower proportions in comparison to milk and faeces and their distribution was unique among most of the individual blood sample pools (Fig 4.21B). Interestingly however, among the 15 most abundant taxa, 13 taxa were observed in common between the three niches except for *Anaplasma* and *Pseudomonas*.

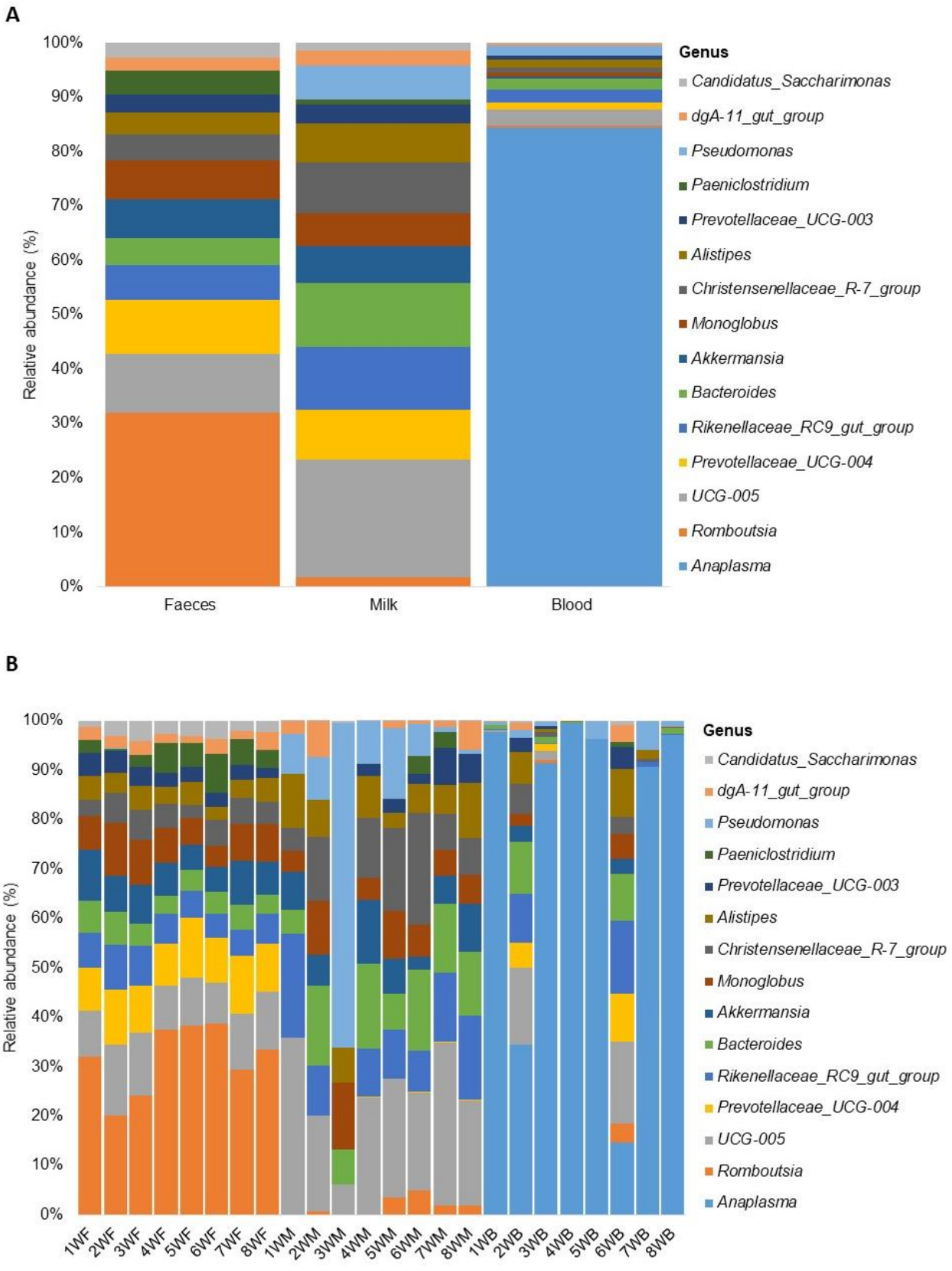

Fig 4.21 A & B: Distribution of genus-level taxa averaged between (A) and within faecal, milk, and blood groups (B).

Importantly among the blood group was the observation of an overrepresentation of one taxon (*i.e*. *Anaplasma*) in the majority of the samples (6/8), with the exception of two samples (*i.e*. 2WB and 6WB) which although were positive, the genus was in much lesser abundance.

It was noted earlier that among blood samples, 2WB and 6WB possessed similar taxa in more or less equal proportions to each other with a moderate proportion of *Anaplasma* in relation to the other blood samples. It can be seen also in Fig 4.21B that these two samples have similar community profiles to those of milk and faeces with regards to the distribution of the 15 most abundant taxa, except for *Anaplasma* which was notably absent in milk and faecal samples.

It could be further noted in the overall comparative analysis that sample 2s of faeces, milk and blood (*i.e*. 2WF, 2WM & 2WB) possessed the same taxa in similar proportions with the 6s (*i.e.* 6WF, 6WM & 6WB) all across the three niches. Interestingly the trio of 2s are made up of samples from animals originating from one kraal while that of 6s are from animals originating from another.

Faecal sample pools seemed to be more consistent with regards to the taxa contained within and their proportions, displaying a balanced profile. Milk somewhat displayed some level of consistency with regards to the taxa, however the proportions in each sample pool were irregular when compared against one another. In contrast, blood samples had the most inconsistent and variable constituents of the taxa when compared to both milk and faeces.

Generally when looking at the relative abundances of the 30 most abundant genus-level taxa across the three sample groups (Fig 4.22): faecal samples had higher abundance values for the taxa contained within; milk on the other hand had more or less evenly distributed or moderate abundance values of taxa overall; while blood recorded the least abundance values of taxa except in the case of *Anaplasma*, which in turn became the most abundant taxon overall despite being absent in faeces and milk. The majority of the remaining taxa (not plotted) however, corresponds to the milk group, corroborating what was observed and discussed above where milk was found to contain the highest number of microbial taxa overall.

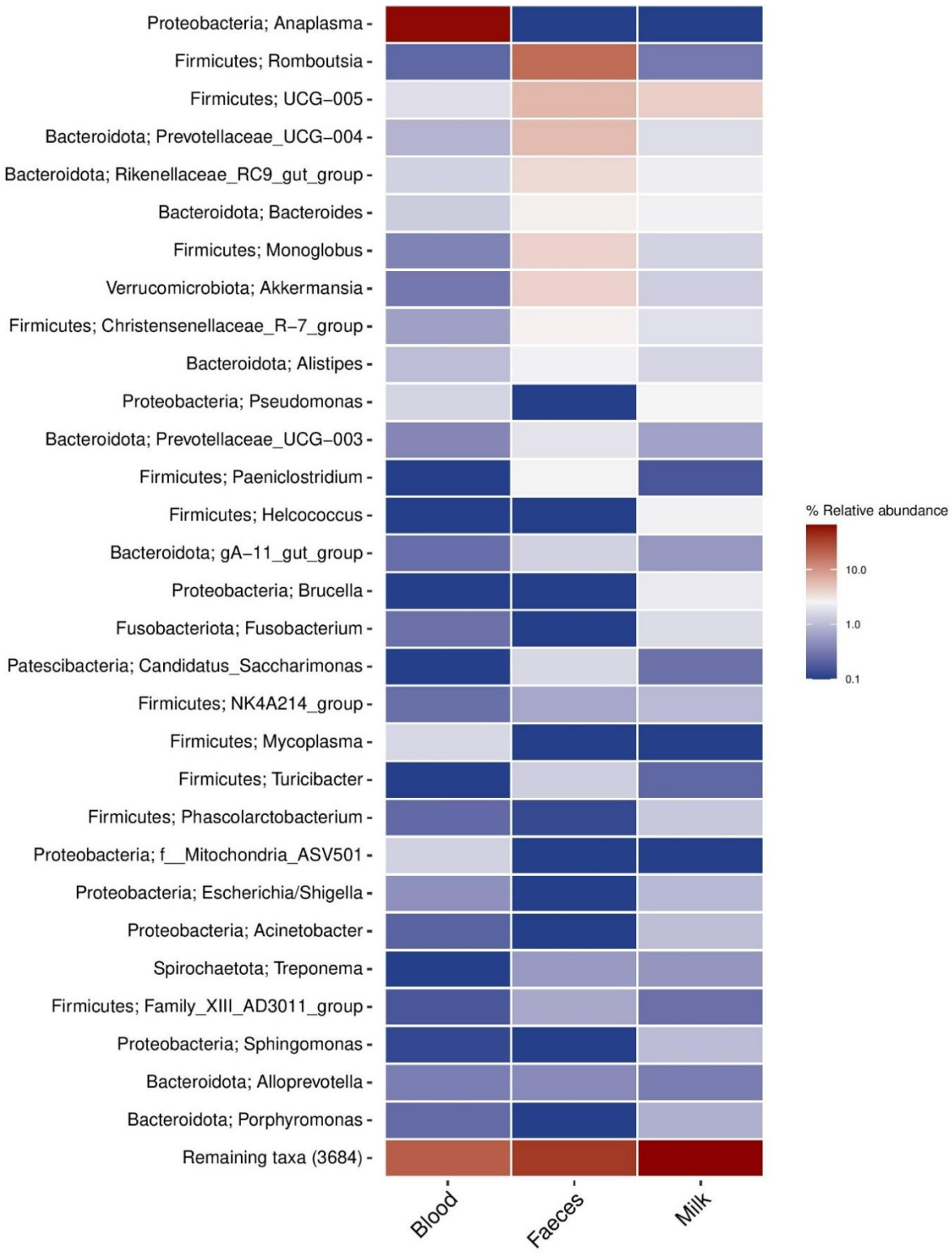

Fig 4.22: Heatmap representing log2 relative abundances of the 30 taxa comprising the most abundant taxa detected amongst the three sample groups (only taxa with relative abundances of ≥ 0.1% are shown for clarity and visualization purposes). The relative abundance values of the bacterial genera within each sample group are depicted by color intensity in the legend indicated

on the right of the figure. Clusters are based on the three sample groups along the x-axis and the taxa are indicated along the y-axis.

A secondary aim of this work was to identify the taxa shared between the three main groups of samples *i.e*. faeces, milk and blood. For this purpose, a comparative analysis of the microbiota detected in each sample group was conducted to determine the extent of overlapping among them using UpSetR intersection plots at family and genus levels (Fig 4.23-4.24; Annexure B).

Of the 243 microbial families detected, there were 49 which were shared between the faecal, milk and blood groups. Milk alone had 158 unique microbial families, while blood had 4 and faeces 2. A total of 26 families were exclusively shared between blood and milk, 3 between faeces and milk and 1 family between faeces and blood.

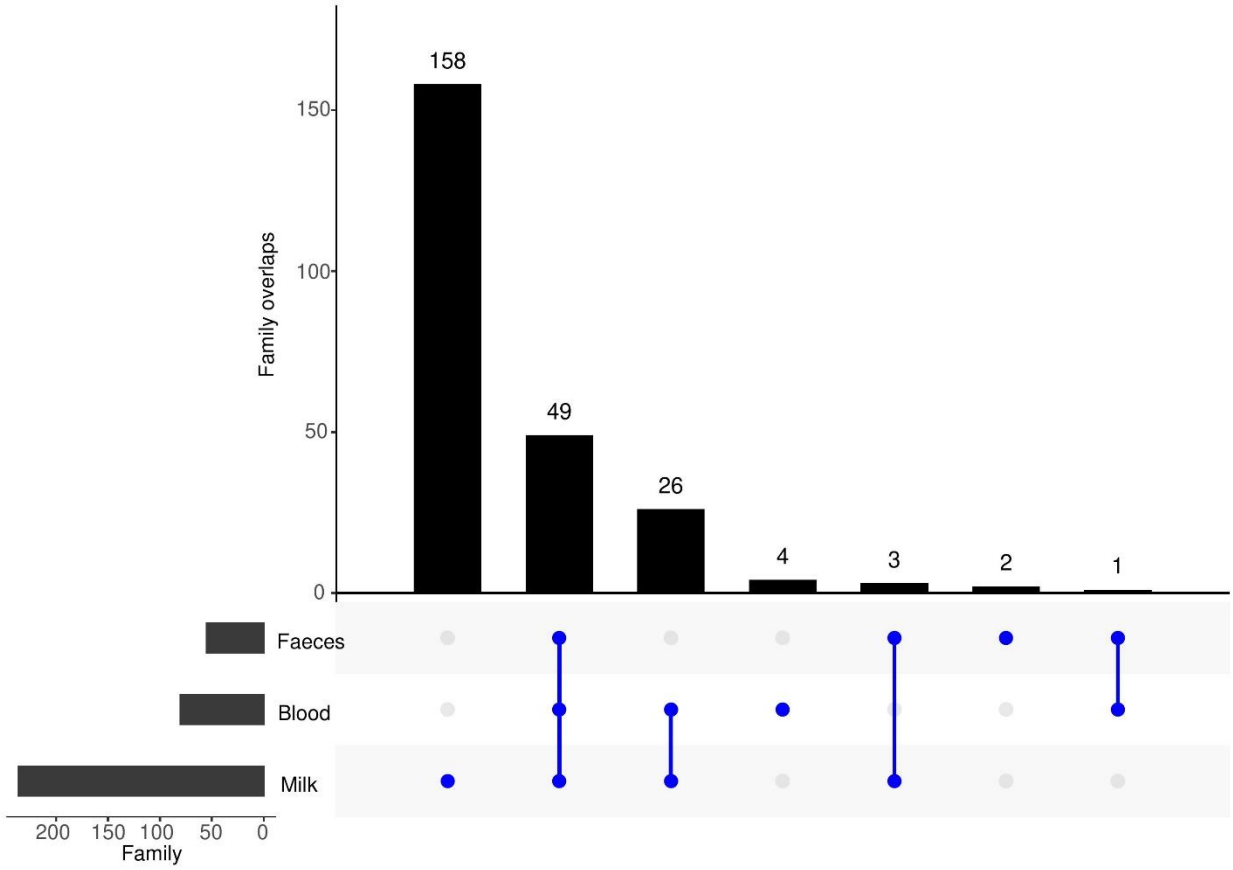

Fig 4.23: UpSetR intersection plot showing number of unique and shared taxa at family level between faeces, milk and blood groups.

Furthermore, from this analysis a total of 58 genus ranking taxa were found to be shared between the three main sample groups accounting for 39.9% of the overall relative abundance as well as 95.3%, 51.4% and 18.6% of the respective relative group abundances of faeces, milk and blood. These taxa and associated classification, abundances and relative abundances are listed in Table 4.15. They were largely dominated by members of Firmicutes (n = 33; 57%) and Bacteroidota (n = 10; 17%).

The shared taxa were represented by multiple ASVs which were mainly exclusively present in either faeces, milk or blood but rarely occurring concurrently across the three niches. Those that did occur across the three niches simultaneously were representatives of 15 genus-level taxa *i.e. Romboutsia* (*e.g.* ASV49, ASV80); *UCG-005* (*e.g.* ASV2034, ASV630); *Prevotellaceae_UCG-004* (*e.g.* ASV2727, ASV2935); *Rikenellaceae_RC9_gut_group* (*e.g.* ASV1961); *Bacteroides* (*e.g.* ASV12295, ASV2471); *Akkermansia* (*e.g.* ASV2594, ASV1343); *Christensenellaceae_R-7_group* (*e.g.* ASV1961); *Alistipes* (*e.g.* ASV1907, ASV7566); *Prevotellaceae_UCG-003* (*e.g.* ASV976, ASV5359)*; dgA-11_gut_group* (*e.g.* ASV1250); *Turicibacter* (*e.g.* ASV10387); *Fusobacterium* (*e.g.* ASV160); *Phascolarctobacterium* (*e.g.* ASV1235, ASV2614); *Coprococcus* (*e.g.* ASV9068) and *Mailhella* (*e.g.* ASV5380).

The ASVs detected concurrently across the three niches mostly occurred in the corresponding groups of samples across the niches (shown in red text on Table 4.15) except for *Akkermansia*, *Alistipes*, *Prevotellaceae_UCG-003*, *dgA-11_gut_group*, *Phascolarctobacterium*, *Coprococcus* and *Mailhella*, which were randomly dispersed across samples between the three niches.

Table 4. 15: Shared taxa between faeces, milk and blood samples with their respective overall raw and relative abundances

Phylum	Class	Order	Family	Genus	Overall raw abundance	Overall relative abundance (%)
Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	*Romboutsia*	26370	7.4
Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	*UCG-005*	15836	4.4
Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotellaceae_UCG-004*	11161	3.1
Bacteroidota	**Bacteroidia**	**Bacteroidales**	**Rikenellaceae**	***Rikenellaceae_RC9_gut_group***	**9665**	**2.7**
Bacteroidota	**Bacteroidia**	**Bacteroidales**	**Bacteroidaceae**	***Bacteroides***	**8197**	**2.3**
Verrucomicrobiota	Verrucomicrobiae	Verrucomicrobiales	Akkermansiaceae	*Akkermansia*	7721	2.2
Firmicutes	Clostridia	Monoglobales	Monoglobaceae	*Monoglobus*	7626	2.1
Firmicutes	Clostridia	Christensenellales	Christensenellaceae	*Christensenellaceae_R-7_group*	6644	1.9
Bacteroidota	**Bacteroidia**	**Bacteroidales**	**Rikenellaceae**	***Alistipes***	**6137**	**1.7**
Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotellaceae_UCG-003*	3979	1.1
Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	*Paeniclostridium*	3703	1.0
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	2957	0.8
Bacteroidota	Bacteroidia	Bacteroidales	Rikenellaceae	*dgA-11_gut_group*	2910	0.8
Patescibacteria	Saccharimonadia	Saccharimonadales	Saccharimonadaceae	*Candidatus_Saccharimonas*	2629	0.7
Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	*NK4A214_group*	2232	0.6
Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	*Turicibacter*	2164	0.6
Fusobacteriota	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	*Fusobacterium*	1796	0.5
Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Anaerovoracaceae	*Family_XIII_AD3011_group*	1503	0.4
Spirochaetota	Spirochaetia	Spirochaetales	Spirochaetaceae	*Treponema*	1468	0.4

Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	*Alloprevotella*	1359	0.4
Firmicutes	Negativicutes	Acidaminococcales	Acidaminococcaceae	*Phascolarctobacterium*	1189	0.3
Firmicutes	Bacilli	Bacillales	Bacillaceae	*Bacillus*	1121	0.3
Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	*UCG-002*	1096	0.3
Firmicutes	Clostridia	Oscillospirales	Butyricicoccaceae	*UCG-009*	1051	0.3
Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	*Candidatus_Soleaferrea*	1043	0.3
Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	*Coprococcus*	1013	0.3
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*	981	0.3
Actinobacteriota	Coriobacteriia	Coriobacteriales	Atopobiaceae	*Olsenella*	845	0.2
Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	*p-1088-a5_gut_group*	828	0.2
Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotellaceae_UCG-001*	820	0.2
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	774	0.2
Bacteroidota	Bacteroidia	Bacteroidales	Marinifilaceae	*Odoribacter*	634	0.2
Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	*Oscillibacter*	518	0.1
Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	*Agathobacter*	488	0.1
Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	*Lachnospiraceae_NK3A20_group*	454	0.1
Desulfobacterota	Desulfovibrionia	Desulfovibrionales	Desulfovibrionaceae	*Mailhella*	435	0.1
Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	*Flavonifractor*	434	0.1
Fibrobacterota	Fibrobacteria	Fibrobacterales	Fibrobacteraceae	*Fibrobacter*	423	0.1
Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	*Ruminococcus*	383	0.1
Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	*Dorea*	305	0.1
Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	*Lachnospiraceae_FCS020_group*	261	0.1
Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	*Acetitomaculum*	253	0.1
Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	*Solibacillus*	233	0.1
Elusimicrobiota	Elusimicrobia	Elusimicrobiales	Elusimicrobiaceae	*Elusimicrobium*	220	0.1

Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Anaerovoracaceae	*Anaerovorax*	213	0.1
Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	*Incertae_Sedis*	213	0.1
Firmicutes	Clostridia	Lachnospirales	Defluviitaleaceae	*Defluviitaleaceae_UCG-011*	200	0.1
Firmicutes	Bacilli	Erysipelotrichales	Erysipelatoclostridiaceae	*UCG-004*	189	0.1
Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	*Lachnoclostridium*	181	0.1
Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	*Pirellula*	163	0.05
Firmicutes	Bacilli	Erysipelotrichales	Erysipelatoclostridiaceae	*Erysipelatoclostridium*	144	0.04
Firmicutes	Clostridia	Clostridia_or	Hungateiclostridiaceae	*Ruminiclostridium*	134	0.04
Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	*Roseburia*	118	0.03
Bacteroidota	Bacteroidia	Bacteroidales	Marinifilaceae	*Sanguibacteroides*	107	0.03
Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotellaceae_Ga6A1_group*	95	0.03
Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	*Enterorhabdus*	77	0.02
Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Anaerovoracaceae	*Family_XIII_UCG-001*	76	0.02
Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	*Lachnospiraceae_UCG-008*	37	0.01

Red text = taxa with ASVs that were concurrently shared across the three niches in the same group of animals. Bold (black and red) text = taxa that formed the shared putative core microbiota present in ≥ 75% of faeces, milk and blood sample pools at ≥ 0.1% relative abundance.

The milk group possessed the highest number of genus-level taxa (374) overall and 254 of these were unique to the group as shown on Fig 4.24. Blood and faeces equally possessed 14 genus-level taxa which were unique to the respective groups. The blood and milk groups shared more taxa at this level (100 genera in common, with 42 exclusively shared between the two groups) than faeces and milk (78 genera in common, with 20 exclusively shared) and faeces and blood (64 genera in common, with 6 exclusively shared). These findings at face value are contradictory to what has been discussed above where the community profiles of milk and faeces seem to be more similar to each other than to blood. However, it is important to note that the above analyses were plotted taking the abundance of dominant taxa (*i.e*. 15 and 30 most abundant) into consideration rather than the entire number of shared taxa between the groups which was subsequently done here onwards.

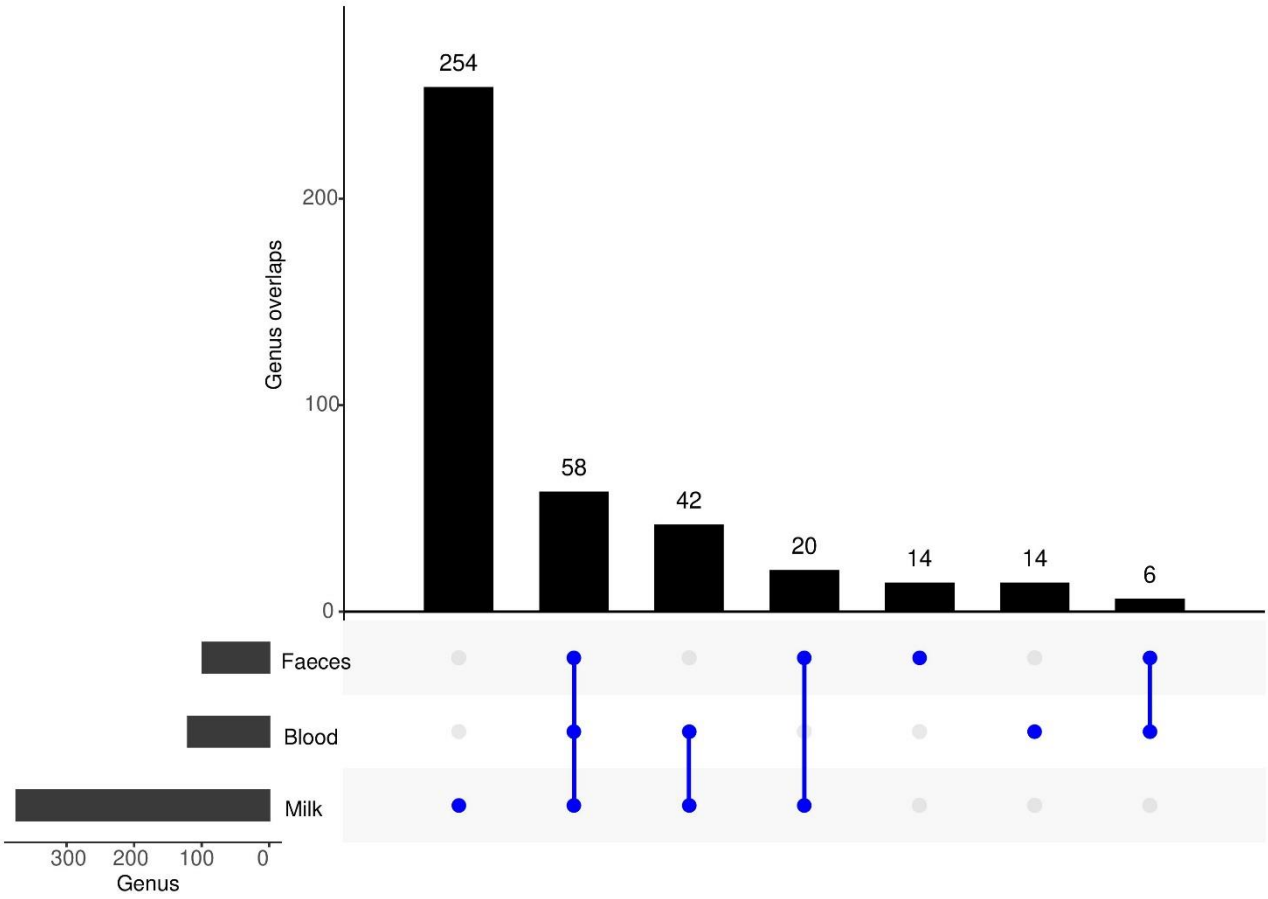

Fig 4.24: UpSetR intersection plot showing number of unique and shared ASVs at genus level between faeces, milk and blood groups.

Among the 58 genus-level taxa shared between the three niches, several important taxa were identified. These include among others *Bacillus, Streptococcus, Akkermansia, Romboutsia, Fusobacterium, Pseudomonas* and *Bacteroides.* At the intersection between blood and milk, important genera such as *Mycoplasma, Escherichia/Shigella, Porphyromonas, Staphylococcus,*

Campylobacter, *Klebsiella* and *Peptostreptococcus* were identified among others. While at the intersection between blood and faeces *Prevotella* could be seen (Annexure B).

Furthermore, to discern the composition of a putative shared core microbiota, the core microbiotas estimated within each sample group (shown in Tables 4.5, 4.8 & 4.11) were compared. Taxa consistently shared between the three sample groups, prevalent in ≥ 75% of the sample pools and detected at ≥ 0.1% relative abundance per group formed the putative shared core microbiota. From this analysis, at the specified criteria, it could be determined that only three (out of 58) taxa at genus level were commonly shared among faeces, milk and blood. These included *Rikenellaceae_RC9_gut_group*, *Bacteroides* and *Alistipes* all under the phylum Bacteroidota, with the respective overall relative abundances of 2.7%, 2.3% and 1.7% (Table 4.15, bold text).

Within group analysis by PCAs revealed that faecal sample pools grouped closely to one another with a few outliers; however, blood and milk samples were scattered across two-dimensional scale. Interestingy, the total variation between samples was greatest among blood samples (52.4%), followed by faecal samples (34.6%), then milk samples (32.4%) as can be seen in Fig 4.5, 4.10 and 4.15.

To compare whole microbial composition between sample groups, β-diversity was calculated using PCoA on weighted UniFrac and Bray distance metrics as well as NMDS analysis on Bray distance metric. The plots were constructed using genus-level taxonomic profiles, with bacterial genus prevalences as covariates and sample type as categorical variables. Generally, the plots showed clear clustering of microbial communities by sample group with a few outliers due to noisy samples (Fig 4.25 - 4.27).

Using the weighted UniFrac distance metric (Fig 4.25), three clusters by sample group could be seen with a total variation of 41.9% between the groups. Although the variation was little, there was an indication that the sample groups contained distinct microbial communities, with milk and faecal samples clustering quite closely to one another.

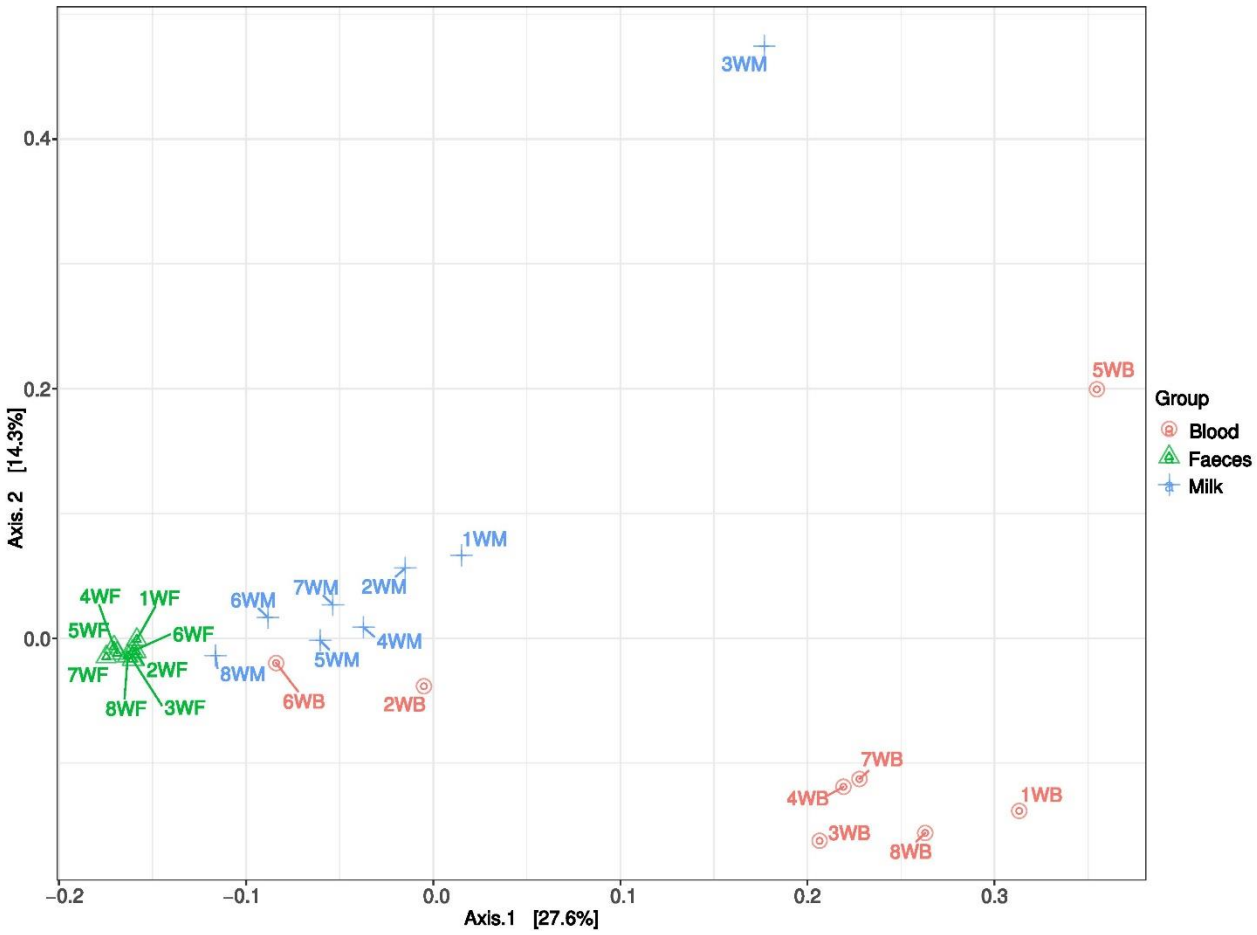

Fig 4.25: β-diversity shown with PCoA of faecal (green), milk (blue) and blood (red) samples based on weighted UniFrac distances calculated using normalized data (log2-fold-change).

Phylogenetically this plot indicates that microbial communities hosted within blood samples have a greater phylogenetic distance from milk and faecal communities with the exception of a few outliers (2WB, 5WB and 6WB). Samples 2WB and 6WB clustered closely with milk and faeces, supporting the observation on Fig 4.21B where the community profiles (based on the 15 most abundant taxa) of these two sample pools looked similar to those contained within milk and faeces. Sample 5WB was highly abundant in one taxon (*Anaplasma*) and contained only four taxa when compared to other blood samples which could explain its deviation from the group cluster. One milk pool (3WM) also formed an outlier, this sample was noisy, lacking a number of the highly abundant taxa present in other milk samples.

This plot also indicates that communities hosted in milk and faeces tend to have a shorter phylogenetic divergence, with shared taxa as could be seen in Fig 4.21 A & B.

PCoA based on Bray distance measure showed similar findings to those calculated using weighted UniFrac distance metric, however this metric provides a measure of community composition differences between samples based on ASV counts, regardless of taxonomic assignment. There was exceptionally clear clustering by sample group on the x-axis with the two principal coordinates accounting for 34% of the total data variability. The data evidenced that the type of sample significantly influenced sample ordination. The differences in community structure between samples were statistically significant (Fig 4.26, PERMANOVA: p = 0.01; F = 4.599). Conversely, this method showed clearer patterns of variation between the sample groups than weighted UniFrac distance metric; indicating that these communities are constituted of bacterial taxa in varying abundances within each sample group.

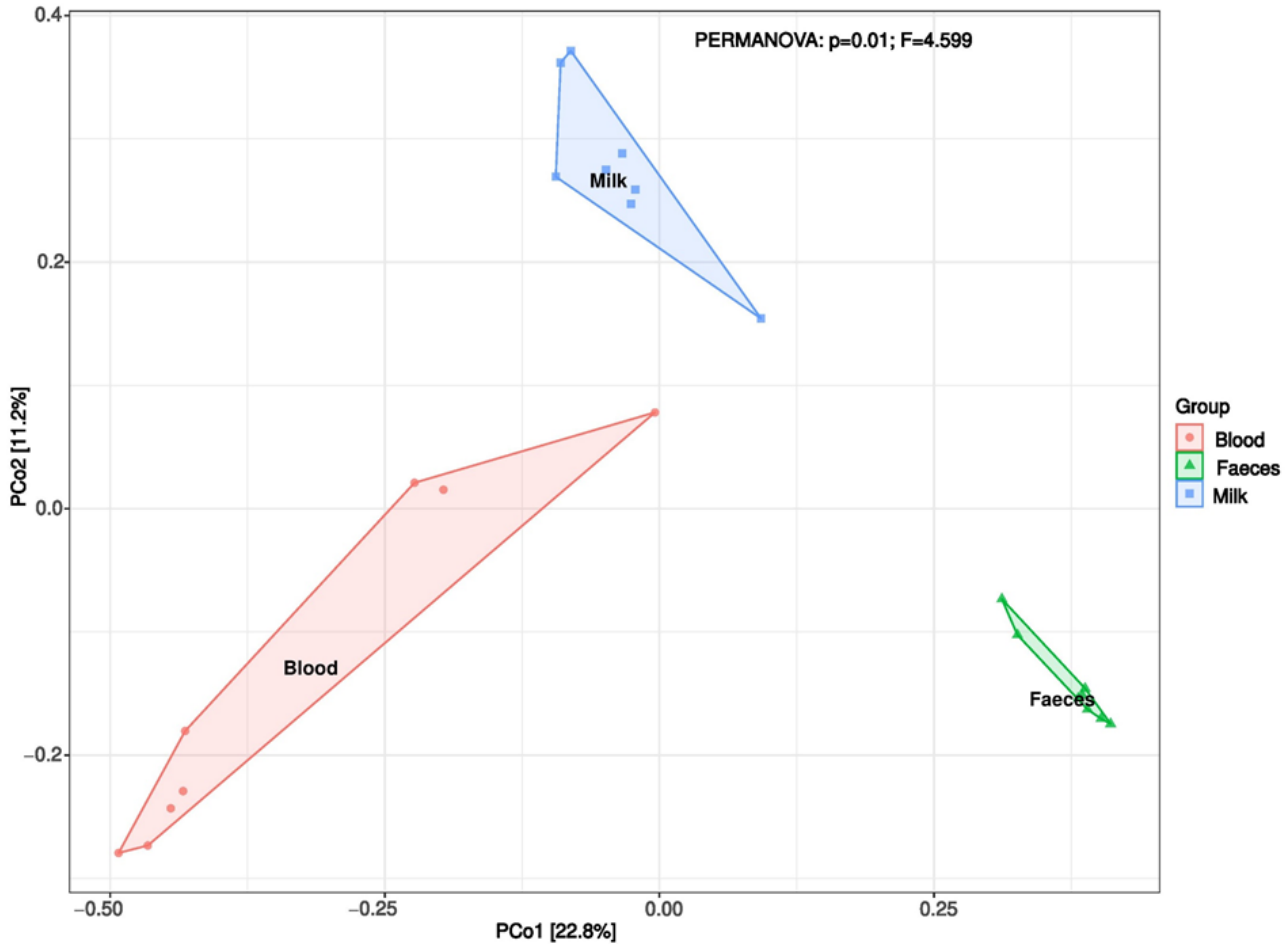

Fig 4.26: β-diversity shown with PCoA of faecal (green), milk (blue) and blood (red) samples based on Bray distance metric, calculated using normalized data (log2-fold-change). The figure shows the relative dissimilarities between microbial communities from different sample groups. Significant differences; $p < 0.05$, PERMANOVA.

Furthermore, a NMDS plot showing the relative dissimilarities between microbial communities from different sample groups was computed on Bray dissimilarity measure. Similarly, there was distinct separation between the three groups and the results were consistent with those of PCoA based on Bray (Fig 4.27; PERMANOVA: p = 0.01; F = 4.599) with a good representation of data (stress = 0.113).

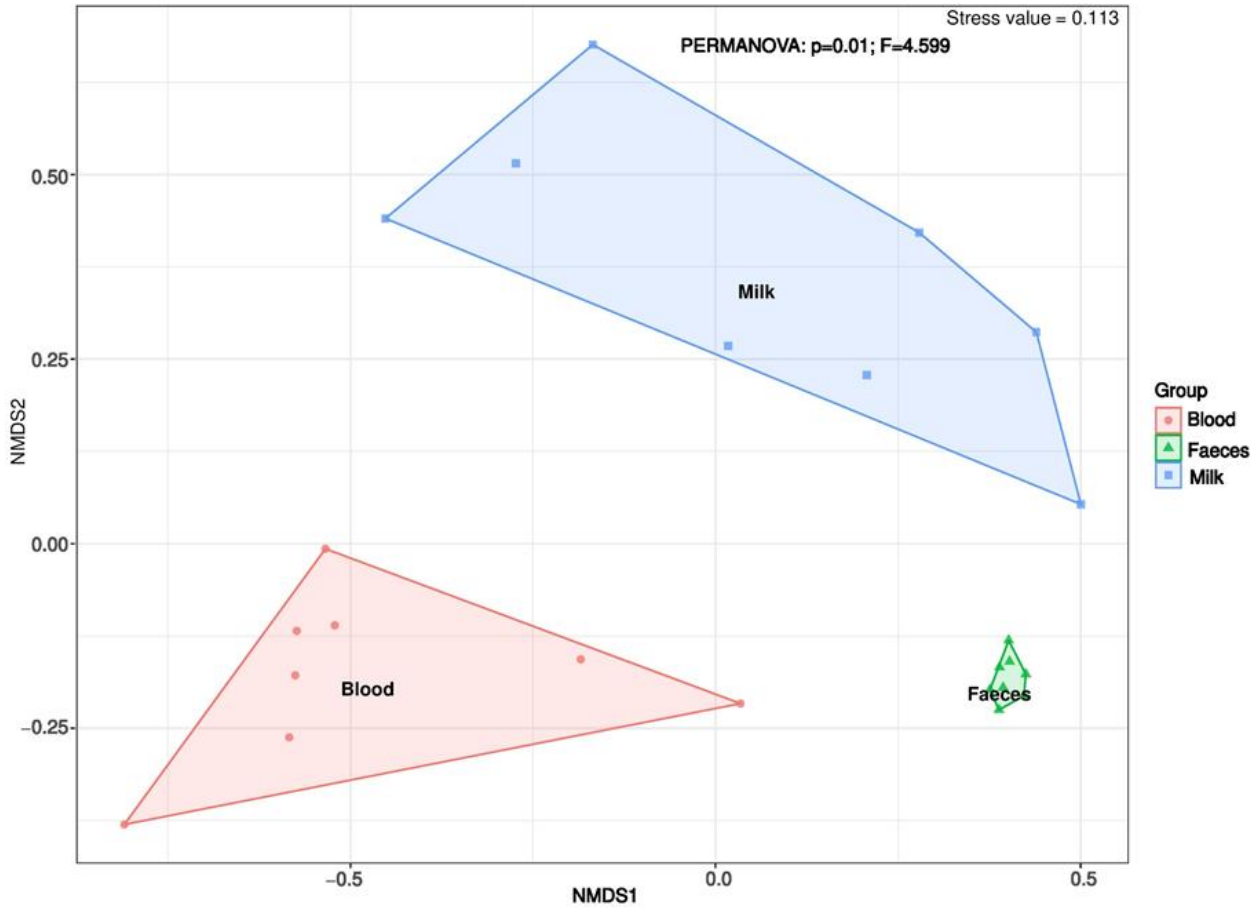

Fig 4.27: β-diversity shown with NMDS plot (stress = 0.113) using Bray dissimilarity metric between sample groups i.e. faecal (green), milk (blue) and blood (red). Significant differences; p < 0.05, PERMANOVA.

In order to determine the compositional differences between the sample groups, analysis of differential abundance on DESeq2 normalized data was performed on the following pairwise combinations: Blood vs Faeces, Blood vs Milk and Faeces vs Milk. Differential abundance testing identified several ASVs that varied significantly in their relative mean proportions with respect to sample groups as determined by the lfcSE pipeline *i.e*. the standard error estimate for the log2-fold-change estimate (*Padj* < 0.01).

When analyzed at the phylum and genus levels, the comparison between Blood vs Faeces, Blood vs Milk, and Milk vs Faeces identified 5 phyla and 18 genera; 4 phyla and 4 genera as well as 4 phyla and 15 genera displaying statistically significant differences, respectively.

Overall 602 ASVs (representing 18 genera) were found to be significantly DA between blood and faecal sample groups. The four most discriminant taxa at genus level are plotted on Fig 4.28. *Anaplasma* (under Proteobacteria phylum) was significantly enriched in blood samples with an average *Pajd* value of 0.0004 (*Padj* < 0.01) and an average log2-fold-change of 6.93. This is line with the finding that *Anaplasma* could only be detected in blood. *Paeniclostridium* (avg *Padj* = 3.94E-05; log2-fold-change = - 8.40), *Romboutsia* (avg *Padj* = 0.0004; log2-fold-change = - 7.78) and *UCG-005* (avg *Padj* = 0.0024; log2-fold-change = - 6.56) under the phylum Firmicutes were greatly reduced in blood than in faeces.

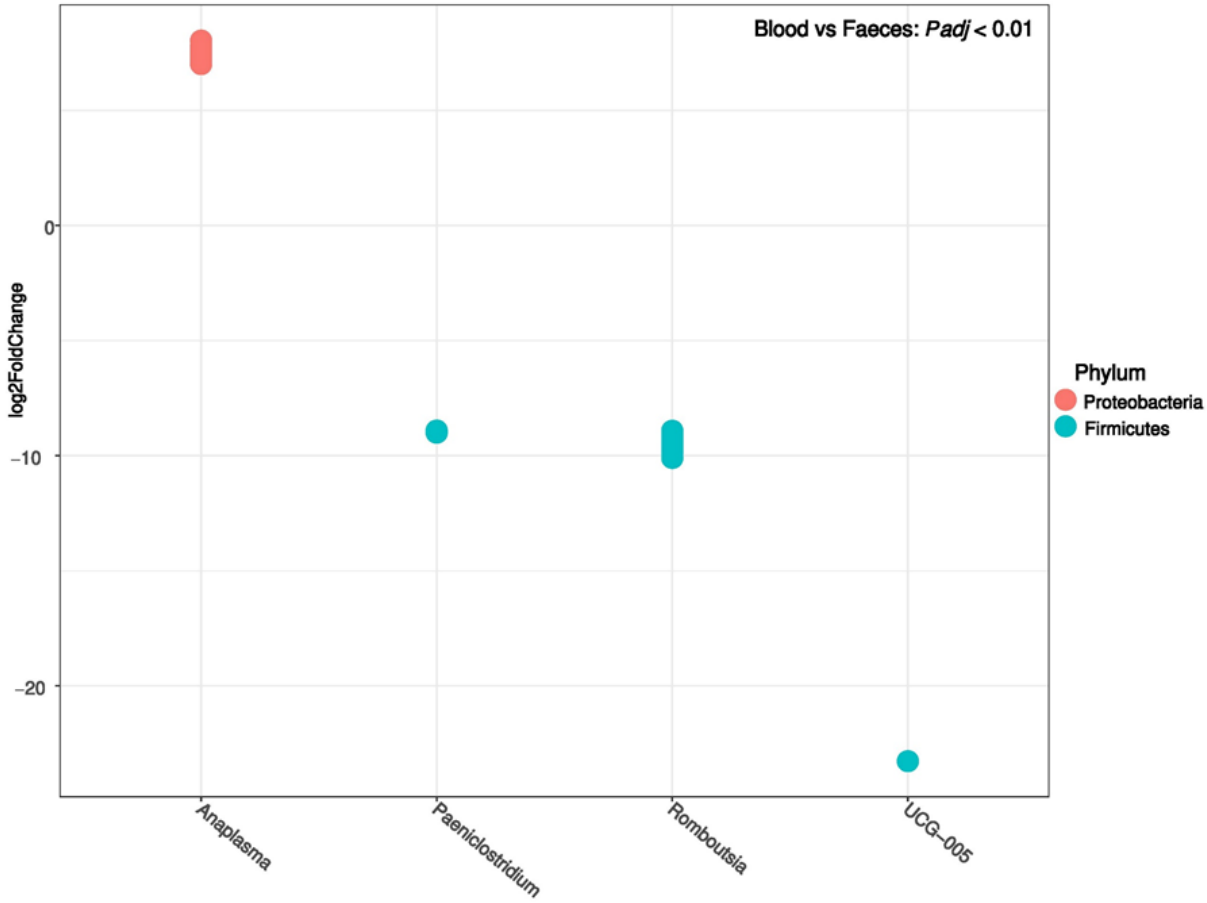

Fig 4.28: Differentially abundant genus-level taxa in blood vs faeces (*Padj* < 0.01). Positive log2-fold change indicates increased abundance in blood compared to faeces, negative log2-fold change indicates decreased abundance. The dots are ASVs representing genus-level taxa.

Differences in abundance between blood and milk groups were defined by 53 ASVs representing four genera which were significantly DA *i.e*. *Anaplasma* (Proteobacteria), *Akkermansia* (Verrucomicrobiota), *Turicibacter* (Firmicutes) and *Prevotellaceae*_UCG-004 (Bacteroidota). *Anaplasma* was significantly enriched in blood (being absent in milk) with an average *Pajd* value of 0.0008 (*Padj* ≤ 0.01; log2-fold-change = 7.01; Fig 4.29), while *Akkermansia* (avg *Padj* = 0.003; log2-fold-change = - 6.88), *Turicibacter* (*Padj* = 0.007; log2-fold-change = - 5.75) and *Prevotellaceae*_UCG-004 (*Padj* = 0.004; log2-fold-change = - 6,47) were significantly reduced in blood than in milk.

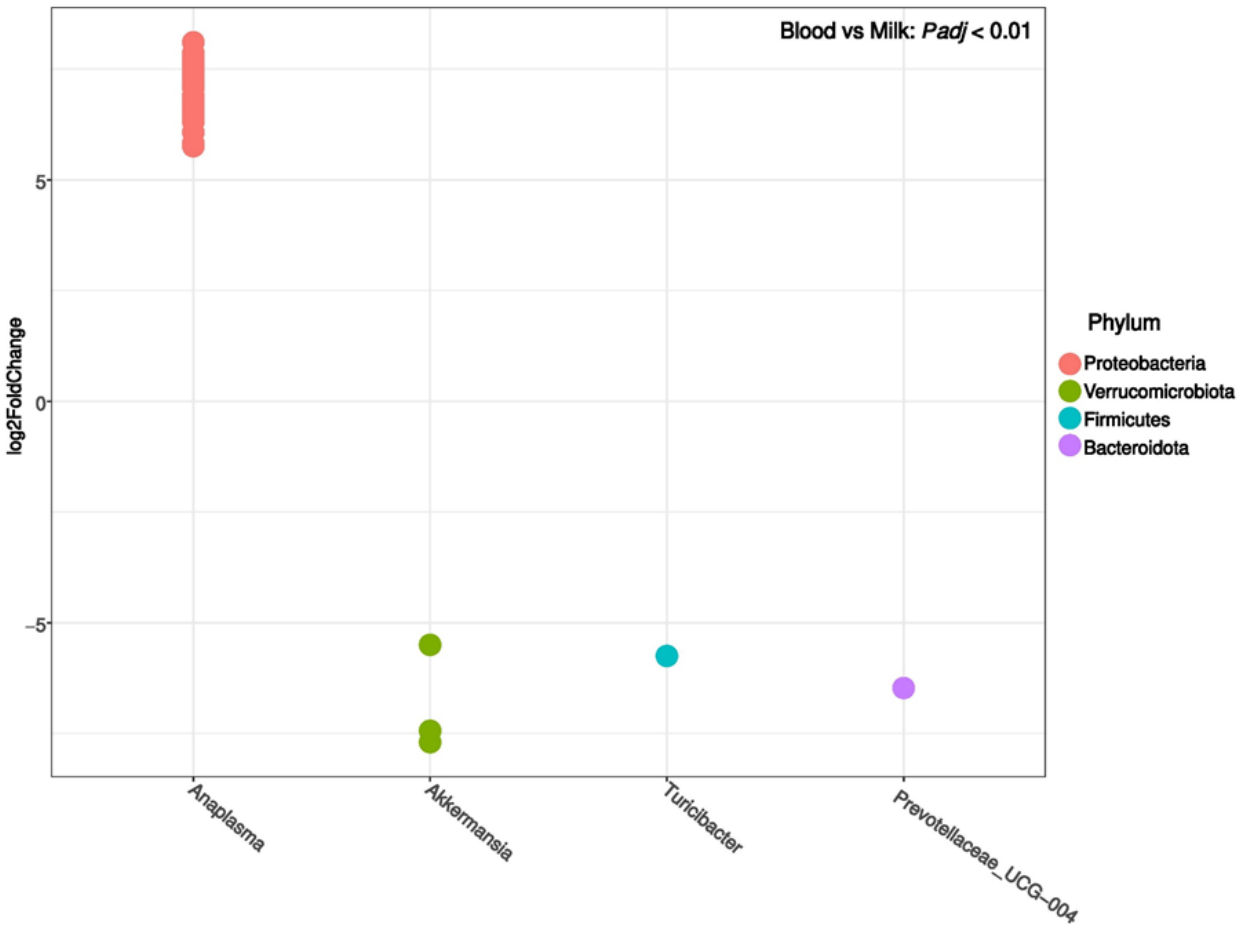

Fig 4.29: Differentially abundant genus-level taxa in bovine blood vs milk (*Padj* < 0.01). Positive log2-fold change indicates increased abundance in blood, negative log2-fold change indicates decreased abundance. The dots are ASVs representing genus-level taxa.

A total of 235 ASVs (representing 15 genera) were DA between faeces and milk. The difference in bacterial abundance between faeces and milk groups was quite interesting with the 9 most discriminant taxa at genus level being all significantly enriched in faeces as opposed to milk (*Padj* < 0.01, Fig 4.30). These taxa included *UCG-005* (avg *Padj* = 0.002; log2-fold-change = 7.76), *Romboutsia* (avg *Padj* = 0.0001; log2-fold-change = 7.43), *Paeniclostridium* (avg *padj* = 0.0006;

log2-fold-change = 7.41), *Candidatus_Saccharimonas* (avg *Padj* = 0.004; log2-fold-change = 5.62), *Prevotellaceae_UCG-004* (avg *Padj* = 0.003; log2-fold-change = 5.19), *Monoglobus* (avg *Padj* = 0.002; log2-fold-change = 6. 08), *Akkermansia* (avg *padj* = 0.003; log2-fold-change = 5. 57), *Turicibacter* (avg *Padj* = 0.0009; log2-fold-change = 5.85) and *Bacteroides* (avg *Padj* = 0.004; log2-fold-change = 4.67). *UCG-005*, *Romboutsia* and *Paeniclostridium* under Firmicutes were particularly overrepresented in faeces. Firmicutes dominated these taxa (5/9), followed by Bacteroidota (2/9).

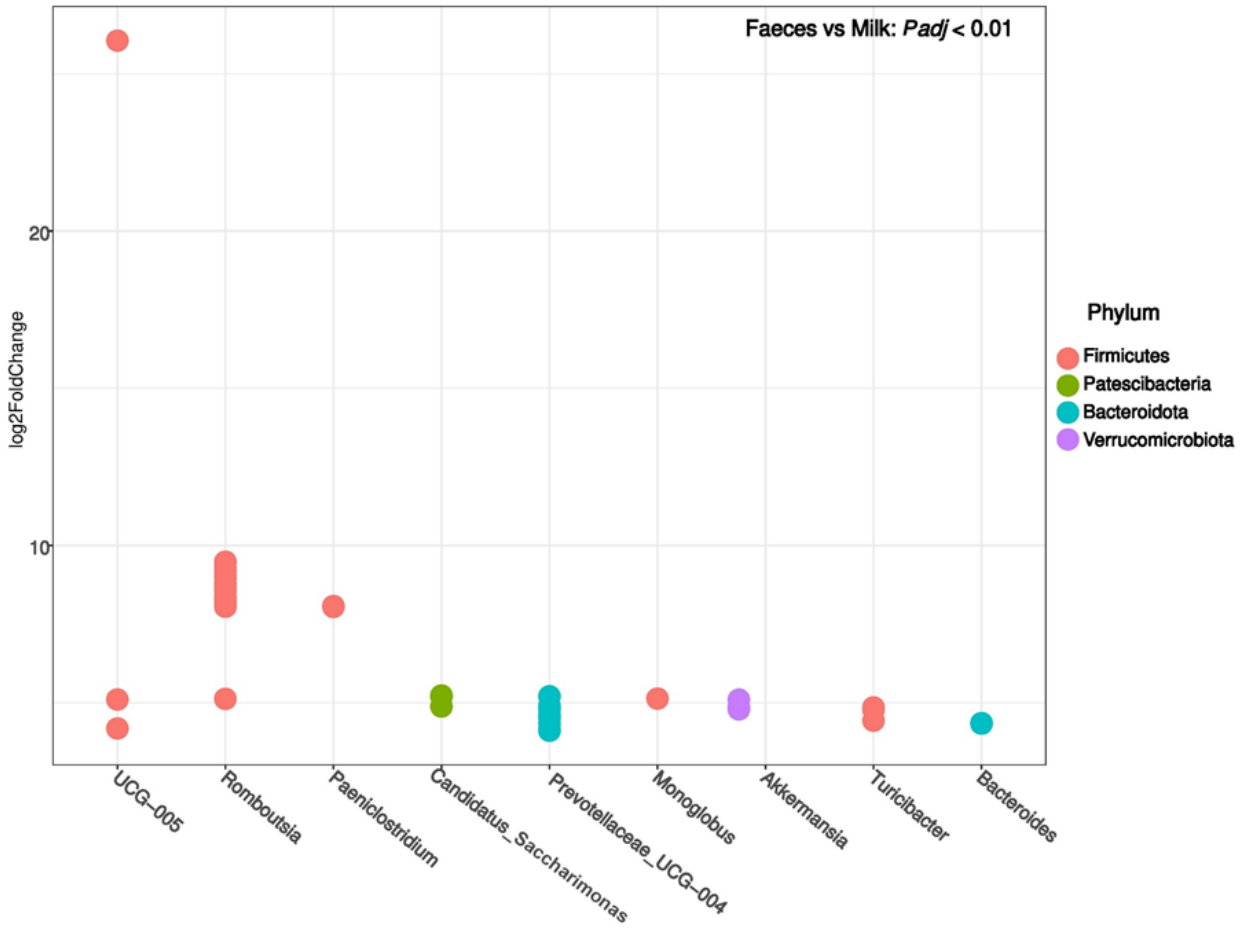

Fig 4.30: Differentially abundant genus-level taxa in bovine faeces vs milk (*Padj* < 0.01). Positive log2-fold change indicates increased abundance of the genera in faeces compared to milk. The dots are ASVs representing genus-level taxa.

4.5.2. Comparative microbial diversity and composition of corresponding faeces, milk, blood and placenta samples

4.5.2.1. Sequence analysis

A comparative analysis between samples 1WF (faeces), 1WM (milk), 1WB (blood) and P (placenta) was conducted in order to determine the correlation between the microbiota contained within the four niches. These four samples belonged to the same animal. Overall the samples yielded a total of 294 923 Illumina reads. A total of 105 962 non-chimeric reads were retained. The least number of reads were in milk (10 470), followed by blood (14 855), faeces (35 490) and the highest number was obtained in placenta (45 147).

4.5.2.2. Microbial diversity analysis

Overall there were 113 genera found across these four samples. The taxonomic ranking per sample type and the overall number of ranks in the analysis between the four sample types is shown in Table 4.16.

Table 4. 16: Microbial taxonomic ranking and overall number of ranks in the analysis between four sample types.

	Kingdom	**Phylum**	**Class**	**Order**	**Family**	**Genus**
Faeces	2	11	15	27	39	56
Milk	3	14	26	47	58	57
Blood	3	6	8	15	21	20
Placenta	1	5	8	10	13	16
Overall number of ranks	**3**	**15**	**30**	**61**	**86**	**113**

*Only kingdom level classification includes members of Archaea and Eukaryota, other downstream analyses were performed strictly on Bacteria

Approximately 99% of the taxa contained in these samples could be assigned at phylum, class, order and family ranks. A total of 99.9% of the taxa were assigned at genus level between the four samples.

The general overview of the microbes contained within each sample type differed starkly in number and proportions as can be seen in Fig 4.31, showing the 15 most abundant genera across the four sample types. An overexpression of one or two taxa in each sample type was observed *i.e.* faeces was dominated by *Romboutsia* (35.1%); milk by *Helcococcus* (14.5%) and *Brucella* (13.7%); blood by *Anaplasma* (94.6%); and placenta by *Fusobacterium* (43.1%).

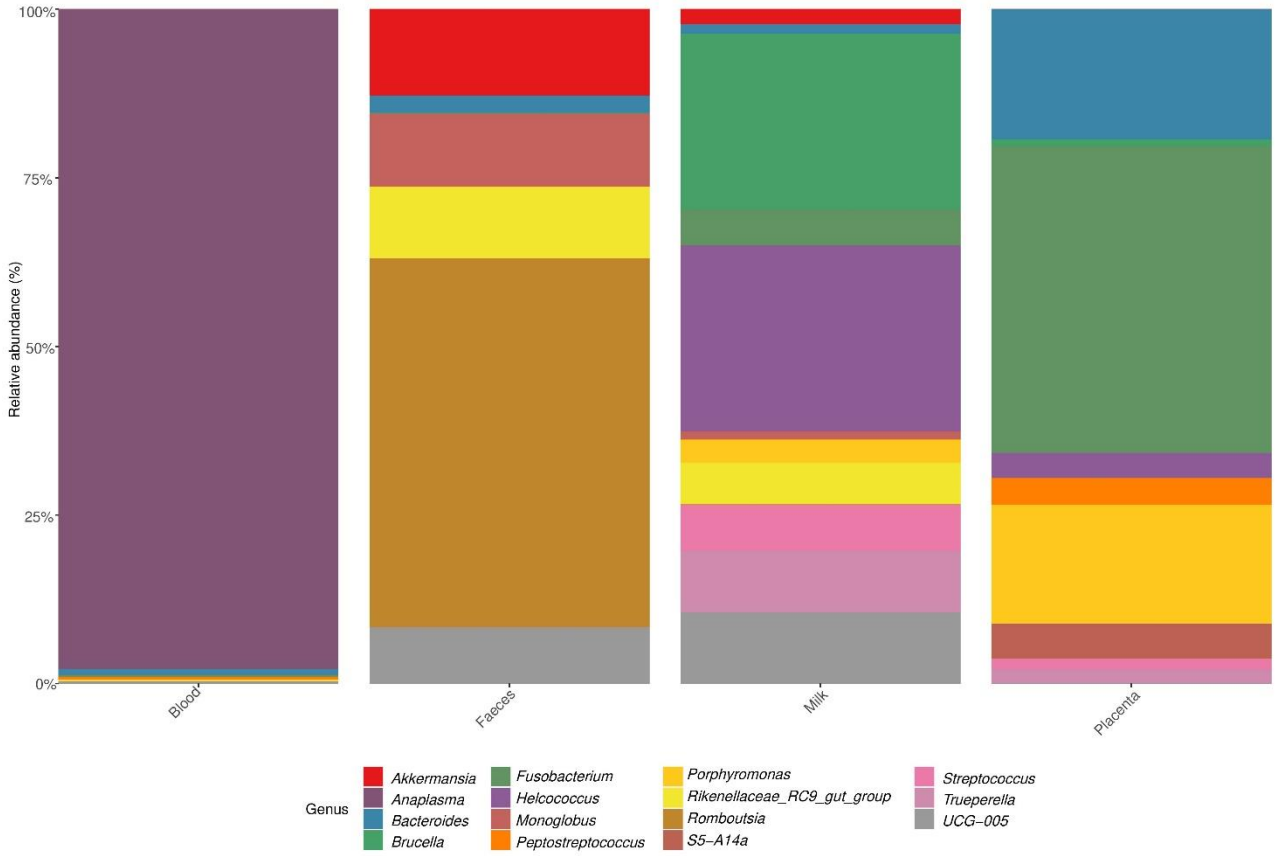

Fig 4.31: Distribution of genus-level taxa across blood, faeces, milk and placenta samples.

This pattern of distribution of taxa is consistent with the observation made in the comparison between the three main sample groups *i.e*. faeces, milk and blood. The number of genus-level taxa that were detected in faeces, milk, blood and placenta were 56 (38), 57 (33), 20 (7) and 16 (7), respectively. The numbers in parentheses (also shown in Fig 4.32) indicate taxa which were

exclusively detected in each sample type. Faeces contained the highest number of exclusively detected genus-level taxa, followed by milk; while blood and placenta were tied.

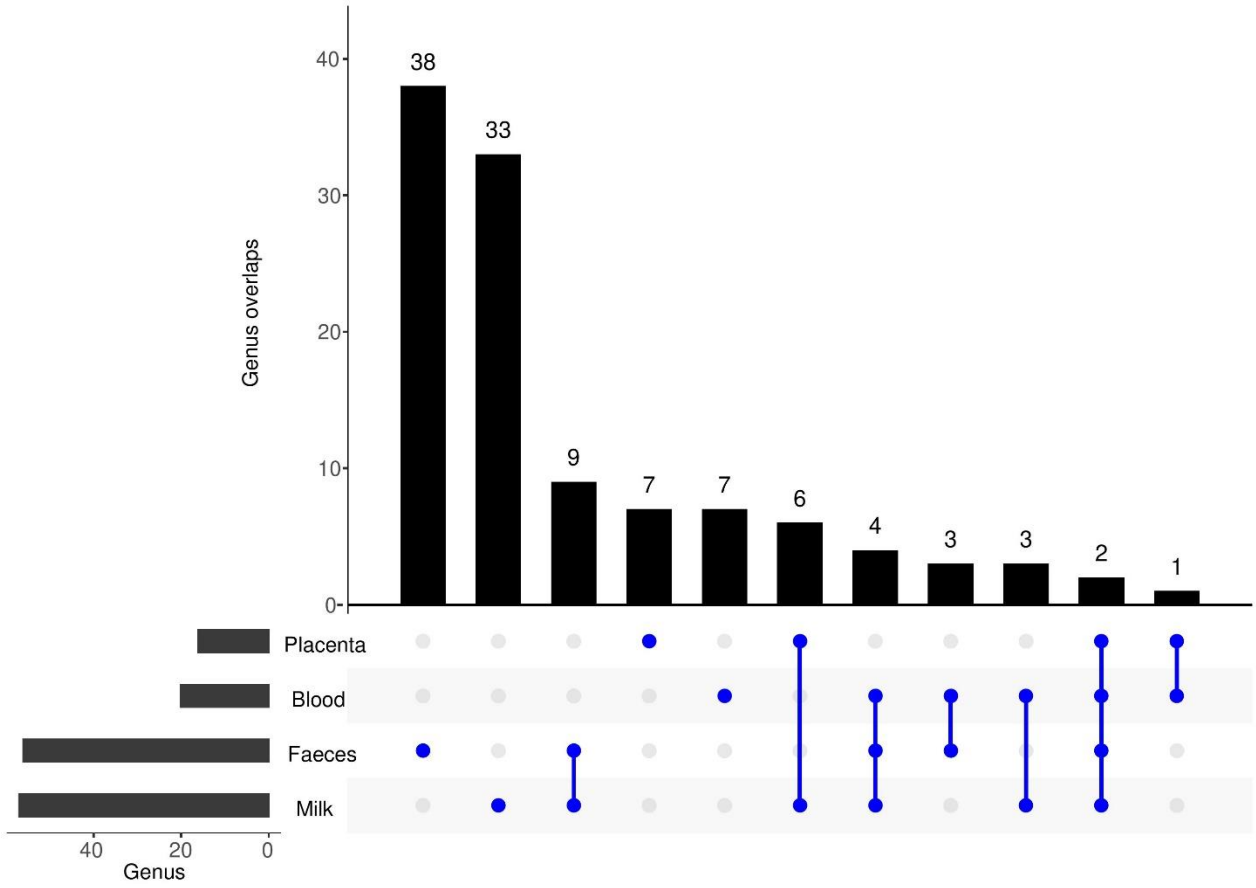

Fig 4.32: UpSetR intersection plot showing number of unique and shared taxa at genus level between faeces, milk, blood and placenta.

Only two genus-level taxa were found to be shared between the four sample types *i.e*. *Fusobacterium* and *Bacteroides*; while between faeces, milk and blood there were six shared genus-level taxa *i.e*. *dgA-11_gut*_group, *UCG-005*, *Rikenellaceae_RC9_gut_group*, *NK4A214_group, Fusobacterium* and *Bacteroides*. Six genus-level taxa were exclusively shared between milk and placenta *i.e*. *Rhodococcus*, *Helcococcus*, *Porphyromonas*, *Streptococcus*, *Brucella* and *Trueperella*. All these genera are considered to be of veterinary significance; *Rhodococcus* and *Brucella* with zoonotic potential. Between faeces and milk samples, nine genus-level taxa were exclusively detected and they included noteworthy microbes such as *Treponema, Akkermansia, Monoglobus* and *Alistipes*. *Pseudomonas, Escherichia/Shigella* and *Acinetobacter* were the only three genera that were found in common between blood and milk.

Seven genera were exclusively detected in blood and these included among others *Anaplasma, Mycoplasma, Bartonella* and *Klebsiella* which are considered to be of veterinary significance. Milk and feaces contained several taxa exclusively, however those of veterinary significance were shared with either blood or placenta as already mentioned.

The microbial diversity of each niche according to Chao1 index was not statistically significant ($p = 0.392$; $p < 0.05$, Kruskal-Wallis), however through this index the highest diversity was observed within faeces (703), followed by placenta (246), milk (189) then blood (116). Faeces as expected had the highest index, milk was replaced by placenta from second place and it ranked third, while blood was the least diverse. This corroborates already discussed findings in prior sections regarding the three main sample groups.

To determine the microbial diversity between the different samples, PCA was conducted using genus-level taxonomic profiles. As shown in Fig 4.33, the samples clustered quite distinctly, with milk and placenta grouping very closely to each other, showing that they were constituted of microbial communities that were similar. The actual reflection however is that milk and faeces shared more taxa (9) than milk and placenta (6). Thus, leading to the speculation that the close clustering was influenced by the distribution of the most abundant taxa shared between the two samples rather than the actual number of shared taxa.

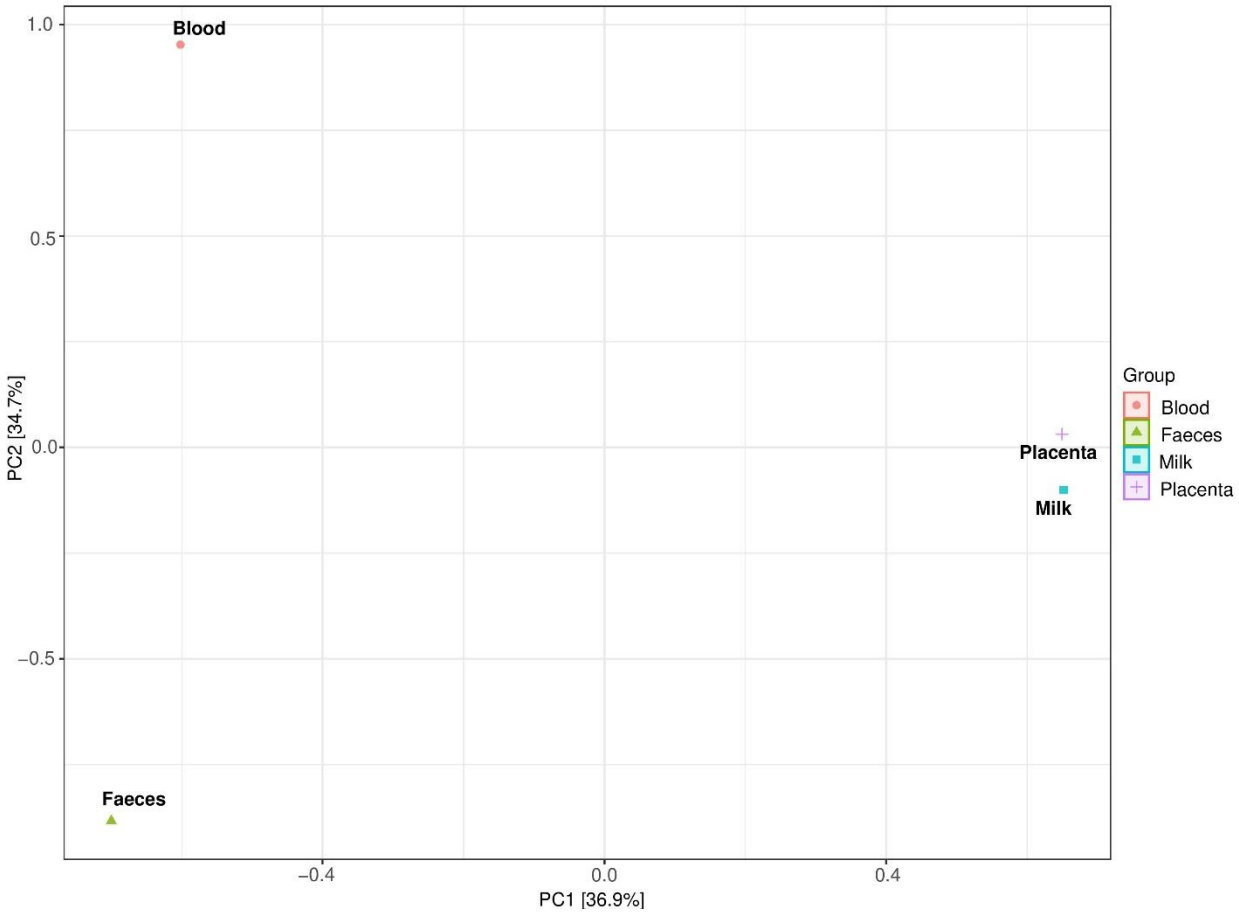

Fig 4.33: β-diversity between faeces, milk, blood and placenta shown with PCA using Hellinger distance metric.

Among the taxa detected within and between the sampled body sites in the entire analysis, there were a number of commonly reported genera of veterinary significance and some with zoonotic potential. These and their respective prevalences per sample group are listed on Table 4.17. The milk and blood seemed to be ideal media to host these potentially pathogenic agents, followed by placenta and the least number were detected in faeces.

Table 4. 17: Prevalence of potentially pathogenic genera of veterinary significance per sample group

Genus	Faeces (%)	Milk (%)	Blood (%)	Placenta (%)
*Anaplasma**	-	-	100	-
*Bacillus**	100	62.5	37.5	-
Bacteroides	100	100	75	100
*Bartonella**	-	-	50	-
*Brucella**	-	25	-	100
*Campylobacter**	-	25	12.5	-
*Ehrlichia**	-	-	12.5	-
*Escherichia/ Shigella**		75	50	100
Fusobacterium	12.5	87.5	37.5	100
Helcococcus	-	62.5	-	100
*Klebsiella**	-	12.5	12.5	-
*Legionella**	-	12.5	-	-
*Mycobacterium**	-	12.5	-	-
Mycoplasma	-	12.5	100	-
Peptoniphilus	-	-	-	100
Peptostreptococcus	-	12.5	12.5	100
Porphyromonas	-	100	25	-
Prevotella	62.5	-	25	-
Pseudomonas	12.5	100	100	-
*Rhodococcus**	-	87.5	12.5	100
*Staphylococcus**	-	50	12.5	-
*Streptococcus**	62.5	62.5	25	100
Trueperella	-	25	-	100
*Corynebacterium**	25	62.5	-	-

Genera containing species with zoonotic potential are marked with an asterisk (*). Negative sign (-) = 0% prevalence

4.6. SELECTIVE SCREENING OF PATHOGENS OF VETERINARY SIGNIFICANCE

From NGS meta-analysis, the single most abundant taxon in the entire analysis was *Anaplasma*. It was significantly enriched in all blood sample pools (1WB - 8WB) with irregularly high ASV counts. Furthermore, high ASV counts of *Brucella* from milk (1WM & 2WM) and placental (P) samples were obtained. Both genera consist of important species of veterinary and medical significance thus, in order to determine their significance among the livestock and potentially the health of the rural community from where the samples were obtained they were characterized to species level. To obtain a general overview of the prevalence of *Anaplasma* among the cattle reared in and around Waaihoek, blood samples from lactating and dry cows were screened via PCR, while for detection of *Brucella* all the milk samples acquired for NGS as well as the placental tissue sample were subjected to culturing and PCR amplification.

4.6.1. Detection and characterization of *Anaplasma* species by PCR

In order to characterize and distinguish between *Anaplasma* species present in the bovine blood, PCR was conducted targeting the bacterial *16S* rRNA gene. A total of n = 110 (1B – 110B) blood samples were screened, *i.e.* n = 88 blood samples from dry cows, together with the n = 22 samples obtained from the lactating cows.

The positive amplicons were verified using confirmed *A. centrale* and *A. marginale* positive controls and representatives sequenced in order to determine the species involved. Genus specific PCR revealed 65% (71/110) prevalence of *Anaplasma*, the positive PCR amplicons can be seen on Fig 4.34 A & B.

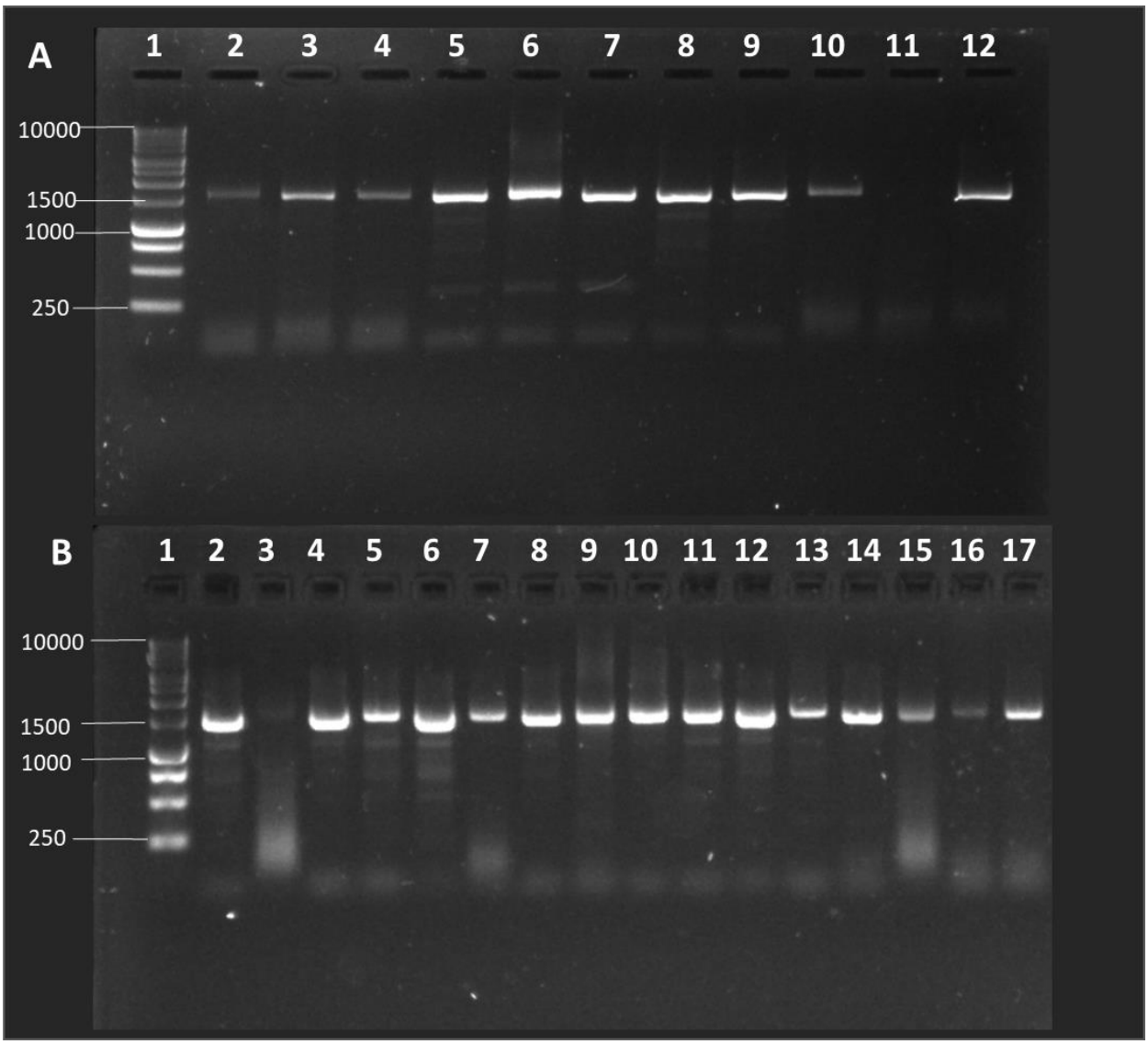

Fig 4.34: Gel electrophoresis of *Anaplasma* PCR targeting the *16S* rRNA gene from blood samples. A: Lane 1 = 1 kb DNA ladder; 2 -10 = positive samples; 11= nuclease free H_2O (-); 12 = *A. marginale* (+). B: Lane 1 = 1 kb DNA ladder; 2, 4-16 = positive samples; 3 = negative sample; 17= *A. centrale* (+).

The PCR positive amplicons were Sanger sequenced for confirmation, 75% of the sequences matched *A. marginale* (Accession numbers: AF414877 and KU686792) at 96.98 – 99.28% identity; while 25% matched *A. centrale* (Accession number: MF289480) at 97.87 – 98.29% identity on the National Center for Biotechnology Information (NCBI) database.

4.6.2. Detection and characterization of *Brucella* species by PCR

All 22 milk samples (labelled 01M - 22M) as well as the placental tissue (P) were subjected to PCR amplification using multiplexed AMOS PCR prior to and after culturing as described in Chapter 3.

Prior to culturing, only the placental sample was positive for *Brucella* and all the milk samples were negative (Fig 4.35 A) however as can be seen on Fig 4.35 B, culturing yielded some positive results. Two known positive controls which were previously verified by Sanger sequencing in another study were used. One positive control (in lane 11) contained *B. abortus* S19 strain (498 bp), while the other (in lane 12) contained *B. canis* (178 bp)

The placental sample yielded an amplicon size of 498 bp on the 1.5% agarose gel (Fig 4.35 B) pre- (lane 8) and post-culturing (lane 9). The obtained DNA sequences matched the insertion sequence *IS711* of *B. abortus* field strain (Accession number: MH615815) at 99.78% identity. AMOS PCR conducted post-culture of the placental tissue produced an additional band around 285 bp as can be seen on the gel (lane 9), however sequencing of this amplicon did not yield any results. Furthermore, post-culture amplification of the milk samples yielded one *Brucella*-positive sample (01M) showing multiple bands equivalent to *B. melitensis* (731 bp) with 100% identity to *B. melitensis* (Accession number: DQ845342); *B. abortus* (498 bp) with 99.32% identity to *B. abortus* (Accession number: MH615815) and the corresponding genus specific *eri* gene fragment at 178 bp on lane 3. All other milk samples remained negative for *Brucella* including 02M, 03M & 04M although these samples constituted a sample pool which was positive via NGS.

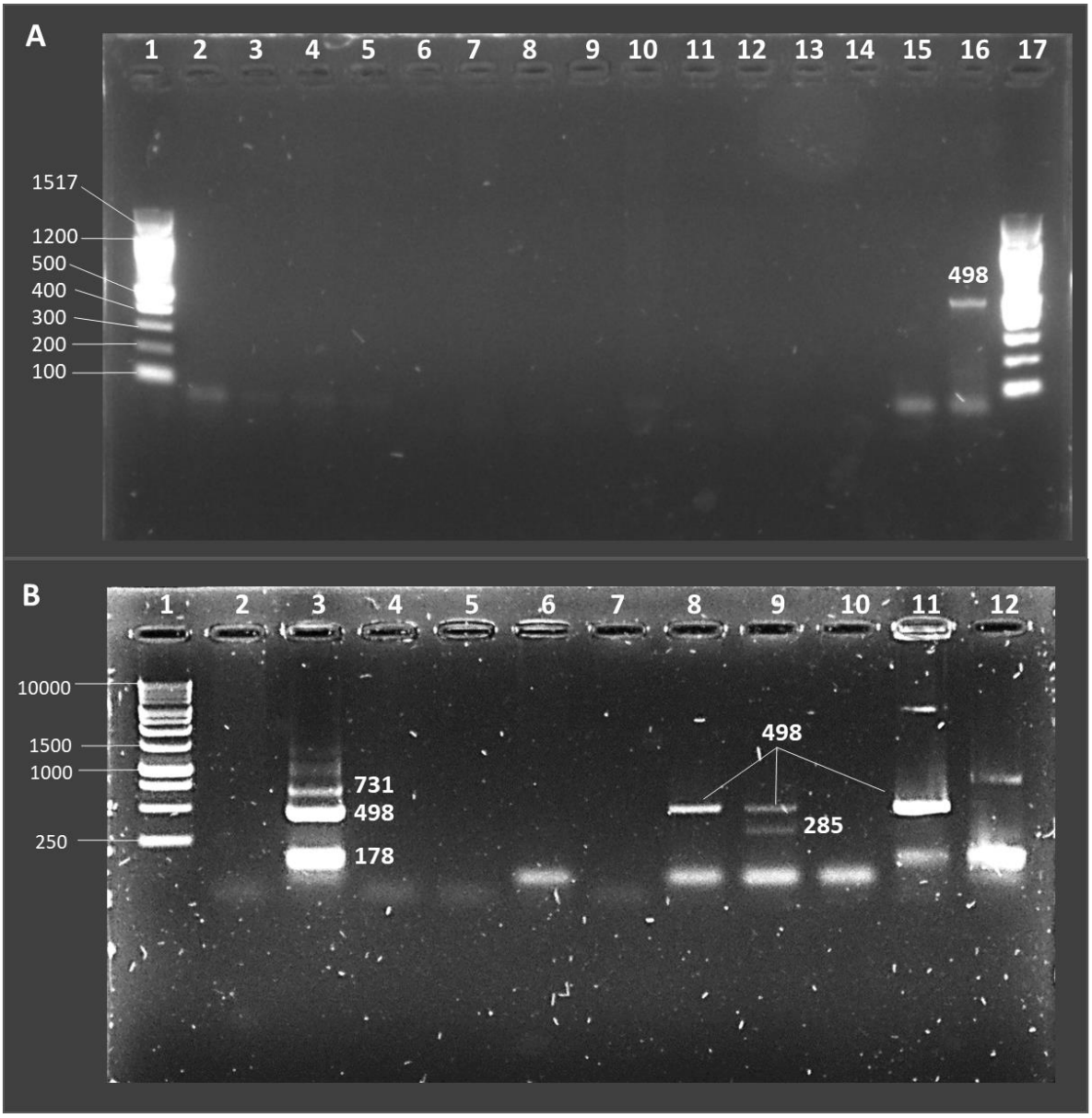

Fig 4.35: Gel electrophoresis of AMOS-PCR conducted on cultured and uncultured milk and placental samples. A: Pre-culture amplification - Lane 1 = 100 bp DNA ladder, 2 - 15 = negative milk samples (01M - 14M), 16 = positive placental sample (P). B: Pre- and post-culture amplification - Lane 1 = 1 kb ladder, 2 = 01M (uncultured); 3 = 01M (cultured); 4 = 02M (uncultured); 5 = 02M (cultured); 6 = 03M (cultured); 7 = 03M (uncultured); 8 = P (uncultured), 9 = P (cultured); 10 = (-) ctrl nuclease free H_20; 11 = (+) ctrl *B. abortus* S19 strain; 12 = (+) *B. canis.*

CHAPTER 5: DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1. DISCUSSION

5.1.1. Sequencing, diversity and taxonomic assignment

In the current study, a total of 602 011 (647 158 including placenta) non-chimeric sequence reads generated from the V3-V4 hyper-variable region were used in the downstream analysis, with an average of 25 084 sequences per sample. Examining the number of reads obtained per sample type showed that they were disproportionate, with faecal samples returning the largest number or reads followed by milk, then blood. This was anticipated due to the difference in the type of samples being analysed, for instance, blood samples generally contain low microbial biomass while faecal samples contain high microbial biomass (Eisenhofer *et al*., 2019). Rarefying to even sampling depth in order to eliminate the sample bias was not an option as large quantities of valid data are lost with this method, interfering with the detection of DA taxa between sampling sites, thus making it inadmissible (McMurdie & Holmes, 2014).

Several taxa were assigned multiple ASVs as the DADA2 sequence inference method has the ability to distinguish sequence variants differing by as little as one nucleotide (Callahan *et al*., 2016; Farrell *et al*., 2019; Kolbe *et al*., 2019). A substantial amount of ASVs were assigned at higher taxonomic ranks but remained unclassified at genus level, while some were unaccounted for at the same level. Previous studies have also reported many animal microbiomes containing certain proportions of unclassified bacteria. The reason for this might due to the limited database of 16S rRNA gene sequences and little research conducted on classification of animal microbiomes (Chen *et al*., 2017), or possibly due to the presence of reads with unclear sequence accuracy (Jeong *et al*., 2021).

Although contaminant sequences were detected in the NTCs, retrospective analysis of contamination proved non-confounding to the findings of this study. The majority of the contaminant microbial genera have been previously detected in negative controls in a minimum of two or more studies and are said to originate from various sources which include kits and reagents contaminated during manufacturing and by commensals on laboratory personnel and equipment (Eisenhofer *et al*., 2019).

The highest count of ASVs were recorded in faeces, followed by milk and then blood. However, the greatest number of genus-level taxa was obtained among milk samples, followed by blood and then faeces. This could be attributed to the fact that although there was a high count of ASVs

in faeces, several clusters were representatives of the same microbial genus and when assembled into taxonomic ranks, they resulted in lower number of taxa as opposed to the observation made in milk and blood. This is in addition to a slightly lower resolution obtained for the faecal group compared to the milk group as mentioned above.

Despite the high resolution power of the pipeline, the 16S rRNA amplicon sequencing depth achieved in the current study was not sufficient for accurate taxonomic assignment at the species level. Although most of the genera had multiple unique ASVs associated with them, it was not possible to determine the species they each uniquely represented apart from only four taxa that were tentatively resolved to species level, representing only a small fraction of the obtained sequences; however, not much could be drawn from these findings due to the scepticism surrounding the accuracy of species level-resolution from sequencing the V3-V4 regions of the 16S rRNA gene where sequences of closely related species have been previously found to be 100% identical (Gupta *et al*., 2019)

Cognisant of the limitations posed by targeting and sequencing of different hypervariable regions as well as the use of different HTS technologies and taxonomic assignment tools, a comparison of the sequence analysis and taxonomic assignment data of the current study is made to that of similar studies below. This is to provide insight into the similarities and differences between the behaviour of microbiota and their distribution patterns among the niches, while drawing as little inferences as possible.

The effect size measurements based on Shannon diversity indices showed that the differences between the sample groups were large enough for assessment of relevant differences between microbial communities represent in the three groups under study. The α-diversity indices of microbial communities contained in faecal samples were higher compared to milk and blood samples (significant at $P < 0.05$). Milk α-diversity values also tended to be higher than those of blood. A study by Young *et al*., (2015) revealed a similar observation where the faecal samples had greater microbial diversity than blood (macrophages) and milk (milk somatic cells). Despite their high α-diversity, faecal samples contained the least number of taxa compared to milk and blood. This could be attributed to the harsh gut environment that possibly does not allow for microbial variety, while blood and milk may be ideal media for cultivation of microbes. The microbial diversity varied significantly between blood and faeces as well as between faeces and milk groups (at $P < 0.05$) indicating that the observed number of ASVs and their abundance between these sample groups were not equally distributed; however, it was not significantly

different between milk and blood. This might have been an indication that the microbes contained within the milk and blood groups were similar in terms of their identities and that there was homogeneity in their abundances. According to Rainard (2017), the similarity between milk and blood microbiota could possibly be explained by the fact that large amounts of blood filter through the mammary glands during lactation, possibly bringing in circulating bacterial components. Alternatively, through the action of patrolling phagocytes in the mammary tissue that occasionally exit the bloodstream, traverse the epithelium, enter the mammary glands and eventually become shed in the milk (Rainard 2017). However, there is no verified microbial translocation pathway between the circulatory system and the mammary glands of ruminants to date thus this remains speculative (Young *et al*., 2015; Rainard, 2017; Derakhshani *et al*., 2018).

The bacterial phyla obtained in faeces, milk and blood (14, 30 & 18, respectively) in the current study were comparable to Young *et al*. (2015)'s findings of 13 bacterial phyla in faeces, 22 in the milk and 15 in blood. Similarly, the most prevalent bacterial groups (in terms of abundance) detected in the faeces and milk included members of the Firmicutes and Bacteroidota phyla; while bacterial sequences from blood were predominantly members of Proteobacteria in both studies. It has been proven that the GIT of calves is seeded before birth with a diverse array of microbiota dominated by Firmicutes, Proteobacteria, Actinobacteriota and Bacteroidota (Alipour *et al*., 2018; O'Hara *et al*., 2020). A commonly reported pattern of GM in descending order of abundance includes Firmicutes, followed by Bacteroidota, then Proteobacteria and / or Actinobacteriota as per observation from literature. This has been corroborated by many authors who conducted research on ruminant GM and reported on the predominance of Firmicutes and Bacteroidota ranging between 63.84 - 68% and 14 - 25.7%, respectively (Oikonomou *et al*. 2013; Liu *et al*., 2014; Mao *et al*., 2015; Young *et al*., 2015; Chen *et al*., 2017; Koester *et al*., 2020). Likewise, this study reports on the predominance of Firmicutes at 64% and Bacteroidota at 25.9% relative abundances, the latter being slightly higher in abundance in comparison to the aforementioned studies. In contrast to these reports however, the Firmicutes and Bacteroidota were succeeded by Veruccomicrobiota (at 4.4%) in this study, following a similar microbial distribution pattern to donkey GM reported by Liu *et al*., (2014) where they obtained an average of 64% relative abundance for Firmicutes, 22% for Bacteroidota and 5% for Verrucomicrobiota. The differences and / or similarities could be explained by diet or different environmental conditions per sampled cohort.

The distribution and proportions of microbial phyla in milk seem to vary per sample group depending on whether the subjects are healthy or mastitic. For instance, in 2018 Derakhshani and colleagues compiled a comprehensive review that described the core MM from clinically healthy cows to be mostly dominated by Firmicutes, Bacteroidota, Proteobacteria and

Actinobacteriota as the main bacterial phyla. The findings of this study resemble the distribution pattern observed by Derakhshani *et al*., (2018), with the following proportions: Firmicutes (39.4%), Bacteroidota (20.4%), Proteobacteria (15.0%) and Actinobacteriota (7.3%). In contrast, Pang *et al*., (2018) reported Proteobacteria as the major phylum ranging from 39.96 - 48.30%, followed by Firmicutes (30.25 - 40.28%), Bacteroidota (8.38 - 12.21%) and Actinobacteriota (5.17 - 11.29%) in milk from both healthy and mastitic quarters. It is difficult to draw conclusions about the findings of this study as mastitis testing was not conducted on the study samples, testing therefore remains imperative in order to properly draw conclusions about the abundance and distribution of the obtained taxa and how they are linked to the health statuses of the sampled animals. However, according to Maity and Ambatipudi, (2021), regardless of whether the mammary gland is healthy or diseased, the main bacterial phyla like Firmicutes, Proteobacteria, Bacteroidetes and Actinobacteria are always there to shape the structure of bovine MM.

The bovine BM obtained in the current study mainly grouped under Proteobacteria (66.4%), Firmicutes (20.6%) and Bacteroidota (9.4%). In contrast, previous studies have reported the predominance of Tenericutes, Proteobacteria and Firmicutes in bovine blood with mean relative abundances of 90.3%, 6.9% and 1.3%, respectively (Jeon *et al*., 2017). On the other hand, Vidal *et al*., (2017) reported on the predominance of Proteobacteria (72.13%), Firmicutes (15.66%) and Bacteroidota (7.81%) in bovine placental samples, in contrast the placental microbiota (PM) in this study consisted of Fusobacteriota (43.2%), Bacteroidota (34.9%) and Firmicutes (15.9%) in descending order of abundance. The differences are possibly influenced by the infection statuses of the animals.

Members of the phyla Firmicutes, Bacteroidota and Proteobacteria, appeared to be the common denominator shaping the microbiota of the studied body sites. Both Firmicutes and Bacteroidota are said to play vital roles in the health of ruminants. Firmicutes function to degrade fiber and cellulose, while Bacteroidota function to degrade carbohydrates and proteins, and facilitate the development of gastrointestinal immunity (Wang *et al*., 2018). Proteobacteria on the other hand are thought to play a key role in preparing the gut of neonates and young animals for colonization by the strict anaerobes required for healthy gut function by consuming oxygen, and lowering redox potential in the gut environment. However, their reputation is often tarnished due to the notoriety of some members being opportunistic pathogenic agents (Moon *et al*., 2018). These roles are mainly associated with the GIT, thus their roles in other niches need further investigation.

The taxonomic profile of microbial sequences identified in the faeces, milk and blood differed between the three niches as shown by the heat map cluster analysis of the most abundant bacterial profiles (Fig 4.22). Dominant taxa (15 and 30 most abundant) in this study grouped faeces and milk samples together mainly based on abundance of taxa, although in actual fact milk and blood samples contained more similar taxa to each other in number. Similarly, in Young *et al*., (2015) bacterial profiles from milk more closely resembled those from faeces based on the hierarchical clustering of the 50 most abundant taxa. It is however unclear how the taxa paralleled between the groups in their overall analysis.

Additionally, in this study the clustering by sample type is shown through β-diversity analysis between samples and groups on ordination plots (PCAs, PCoA and NMDS), statistically significant at $P < 0.05$. The ordination plots enabled easy visual inspection of sample groupings, driven by sample microbial content similarity and / or dissimilarity, abundance and sample type. The PCoA plot indicated that microbial communities hosted within blood samples had a greater phylogenetic distance from milk and faecal communities with the exception of a few outliers. It further suggested that blood samples had a uniquely distinct microbial community compared to the other sample types analysed and that communities hosted in milk and faeces tended to have a shorter phylogenetic divergence with similar taxa and associated abundances. The most dominant taxa (in terms of abundance) influenced ordination on PCoA plots using weighted uniFrac metric despite the true picture of the overall number of taxa contained within and shared between samples. That is to say, even though some sample pools (within group comparison) and sample types (between group comparison) may have shared a higher number of taxa, it did not influence their grouping in two-dimensional space as much as the respective abundances of the taxa contained within. Additionally, PCA (on Hellinger distance metric) and PCoA (on weighted uniFrac distance metric) showed that they were sensitive to noise, *i.e.* samples containing unique ASVs in little abundances as well as those that contained fewer taxa in comparison to the rest. While NMDS and PCoA on Bray distance metric were not affected by noise. The data obtained from PCoA and NMDS plots evidenced that sample type significantly influenced sample ordination, while farm origin and pooling strategy did not have any evident effect on sample clustering.

Moreover, between group analysis on PCoA and NMDS plots were in agreement with the findings of heat map cluster analysis and stacked bar plots. They revealed that faecal samples displayed a more balanced profile (homogenous) with little variation, this was also supported by the composition of the core faecal microbiota. Milk and blood microbiota displayed slightly higher inter group variation with blood samples being distinctly dominated by one genus. The high inter group

variation and much lesser number of species forming part of the core microbiota of milk and blood samples might be an indication that there was no typical milk and blood microbiota. It could be speculated that the source of variation among blood samples is linked to the infection statuses of the animals, particularly in the case of arthropod-borne pathogens observed. Among faecal samples it could be related to diet and the composition of the ruminal microbiota which ultimately shapes the faecal microbiota as previously suggested (Fomenky *et al*., 2018; Cendron *et al*., 2020), while among milk samples, it could perhaps be related to the stage of lactation (Young *et al*., 2015; Zhang *et al*., 2015). There are however a variety of other factors (intrinsic and extrinsic) that could contribute to the observed variation between samples (Deng *et al*., 2019). These may include exposure to microbes from exogenous sources such as bedding material, herd faeces and food sources (Zhang *et al*., 2015; Derakhshani *et al*., 2018). Furthermore, the infection status of the animals is generally known to influence microbial dysbiosis in body sites favouring some microbes (pathogenic) at the expense of others (commensalistic), (Ong *et al*., 2021). This variation could have also arisen simply due to chance as previously suggested (Young *et al*., 2015).

It is not possible to discuss all the taxa obtained in the entire analysis individually. Thus hereinafter, predominant taxa, particularly those which formed part of the core microbiota (at ≥ 0.1% relative abundance and ≥ 75% prevalence) of the different niches as well as those which were significantly DA through discriminant analyses are discussed. Furthermore, the discussion will highlight taxa which although not forming part of the core microbiota, were perceived to be of veterinary and / or medical significance.

5.1.2. Analysis of predominant taxa per sample group

The observed faecal microbiota represents a mixture of taxa containing known anaerobic gut microbes *e.g.* members of genera such as *Clostridium_sensu_stricto_1*, *Romboutsia* and *Bacteroides* (Dowd *et al*., 2008; Cendron *et al*., 2020); typical initial gut colonizers or bacteria found in the intestine but more typically present on other mucosae *e.g. Streptococcus* and *Staphylococcus* (Alipour *et al*., 2018); and bacterial genera with potential health effects on the cattle hosts and their owners *e.g. Bacillus* and *Clostridium_sensu_stricto_1* (Dowd *et al*., 2008).

Of the top ten most abundant genus-level taxa obtained among faecal samples in this study, 6 (60%) of these in no particular order are similar to the findings recorded by Chen *et al.* (2017) from the golden takin (gnu goat) GM, *i.e. Akkermansia, Bacteroides, Alistipes*, *UCG-005, Rikenellaceae_RC9_gut_group* and *Christensenellaceae_R-7_group*. These all formed part of

the core faecal microbiota obtained in this study. Similarly, Dowd *et al*., (2008) also reported the aforementioned bacteria to predominate 50 - 100% of the cattle faecal samples in their study and additionally reported detection of *Prevotella*, *Treponema, Clostridium_sensu_stricto_1*, *Ruminococcus, Streptococcus* and *Roseburia* with a similar prevalence. The additional taxa were also detected in this study ranging between 12.5 - 100% in prevalence, but only *Treponema* and *Clostridium_sensu_stricto_1* formed part of the faecal core microbiota. In addition to the already mentioned *Prevotella* and *Treponema*, Mao *et al*., (2015) also identified *Acetitomaculum, Mogibacterium* and *Acinetobacter* among the predominant genera in cattle GITs. Again, these were detected in the current study, but only *Mogibacterium* formed part of the core faecal microbiota in the current study. There were a number of taxa that formed the core faecal microbiota (43/98; 44%) in this study and they've been reported to be variably present in different portions of the GIT and in faeces by various authors (Dowd *et al*., 2008; Alipour *et al*., 2018; Koester *et al*., 2020). This suggests that there might be a core microbial community that is ubiquitously present within the gut of cattle, eventually shed in faeces and its structure is possibly dependent on individual farm management and husbandary practices.

The milk samples generally consisted of a diverse range of opportunistic and commensal bacteria, possibly inhabiting the teat canal or mammary gland. These included frequently identified bacterial groups across the udder such as lactic acid bacteria *e.g. Lactobacillus* and *Enterococcus*; psychrotrophic bacteria *e.g. Acinetobacter* and *Pseudomonas*; skin-associated bacteria *e.g. Staphylococcus* and *Corynebacterium* and; gut-associated bacteria *e.g. Clostridium_sensu stricto 1* and *Bacillus* (Derakhshani *et al*. 2018; Khasapane *et al*., 2020; Maity & Ambatipudi, 2021). Lactic acid bacteria reportedly predominate bovine milk prior to pasteurization (Maity & Ambatipudi, 2021), while psychrotrophic bacteria successfully establish themselves during cold chain storage (de Oliveira *et al*., 2015; Maity & Ambatipudi, 2021). Taking into consideration the relatively simple nutritional requirements of psychrotrophic bacteria and their ability to survive very low temperatures (de Oliveira *et al*., 2015), it is no surprise that they were dominant in highly nutrional milk samples. The skin- and gut-associated bacteria could possibly be contaminants from stripping of the teats during milk collection and from gut microbes colonizing the teat canal from herd faeces, respectively.

Furthermore, there seems to be a common core milk microbiome represented by particular species which are frequently reported across HTS microbial investigations of healthy, subclinical and clinical mastitic cows, averaging around 20 taxa (Bhatt *et al*., 2012; Oikonomou *et al*., 2012 & 2014; Quigley *et al*., 2013; Young *et al*., 2015; Falentin *et al*., 2016; Taponen *et al*., 2019; Hoque *et al*., 2019). Similarly, in this study the core milk microbiota was constituted of 23 (6.1%) commonly reported taxa. Prevalent taxa (those with a relative abundance of ≥ 1.0% at genus

level) found in the current study, which were also among the most abundant taxa in other studies documenting healthy milk microbiota include: *Pseudomonas* and *Sphingomonas* (Kuehn *et al*., 2013); *Bacteroides, Porphyromonas* and *Fusobacterium* (Oikonomou *et al*., 2014); as well as *Helcococcus* and *Trueperella* (Zhang *et al*., 2015). Other taxa prevalently detected in healthy cows which were present but less abundant (at < 1.0% relative abundance) in the current study were *Corynebacterium, Stenotrophomonas, Streptococcus, Staphylococcus, Bradyrhizobium, Lactobacillus, Comamonas, Chryseobacterium, Brevundimonas, Leucobacter, Butyrivibrio, Facklamia, Mogibacterium, Ruminococcus, Ralstonia, Sphingobium* and members of the family Lachnospiraceae (Kuehn *et al*., 2013; Oikonomou *et al*., 2014; Zhang *et al*., 2015; Addis *et al*., 2016; Derakhshani *et al*., 2018). A number of taxa which are responsible for environmental and contagious mastitis were also detected.The most prevalent taxa that have been detected in milk from mastitic quarters (clinical and subclinical) through HTS based studies, also identified in the current study, include *Trueperella*, *Escherichia/Shigella*, *Streptococcus* and *Klebsiella* (Addis *et al*., 2016; Motaung *et al*., 2017). The former two genera were among the most abundant taxa, while the latter two genera were among the less abundant taxa in the current study. This data is suggestive of the presence of subclinical mastitis in the sampled cows as clinical signs associated with it were not observed. The milk microbiota obtained in this study differs from that which was previously reported from raw milk in South Africa (Free State) by Khasapane *et al*., (2020). In the said study, the authors reported on the predominance of *Clostridium* followed by *Romboutsia, Turicibacter, Dubosiella, Facklamia, Lactobacillus* and *Aerococcus* which although mostly detected in this study (except for *Dubosiella*) were not among the most abundant taxa (ranging between 0.1 - 0.6% among the milk samples). The deviation of results could be expected since as stated by Zhang *et al*., (2015) the teat surface is considered to be the main source of milk contamination, and the composition of the microbial community on the teat surface varies qualitatively and quantitatively from one farm to another.

Of the few bovine blood HTS based studies that exist, *Anaplasma, Mycoplasma*, *Bartonella*, *Pseudomonas, Stenotrophomonas* and *Bacteroides* have been recorded among the most abundant taxa (Young *et al*., 2015; Jeon *et al*., 2017; Kolo *et al*., 2020). These were also detected in the current study and in addition to them, other taxa of interest were detected including *Escherichia/Shigella, Ehrlichia, Fusobacterium, Streptococcus, Prevotella, Rhodococcus* and *Klebsiella* in lower abundances. The majority of the observed taxa were atopobiotic in the blood having possibly entered the bloodstream from their usual sites of colonization such as the gut (*e.g. Rikenellaceae_RC9_gut_group* and *Prevotella*), teats (*e.g. Streptococcus* and *Klebsiella*) and uterus (*e.g*. *Fusobacterium* and *Bacteroides*). They may have been translocated endogenously via some haematogenous mechanism into the bloodstream, however in the case

of arthropod-borne pathogens (*e.g*. *Ehrlichia* and *Bartonella*), these may have been inoculated into the bloodstream from an external source.

Taxa that were detected from the placental sample were identical to frequently isolated uterine pathogens from other studies. HTS based investigations of placental microbiota have previously reported on the predominance of *Trueperella* (*T. pyogenes*)*, Escherichia* (*E. coli*), *Fusobacterium* (*F. necrophorum*), *Porphyromonas* (*Po. levii*), *Prevotella* (*Pr. melaninogenica*), *Helcococcus* (*H. ovis*), *Bacteroides* (*B*. *pyogenes*), *Streptococcus* spp., *Staphylococcus* spp. and *Bacillus* spp. (Jeon *et al*., 2017; Vidal *et al*., 2017; Galvao *et al*., 2019; Lima, 2020, Ong *et al*., 2021). Likewise, from the placental sample, the majority of these genera were detected in high abundances, except for *Prevotella*, *Staphylococcus*, *Escherichia* and *Bacillus* which were not detected. Previous studies indicate that α-hemolytic streptococci, *Staphylococcus* spp., *Bacillus* spp and *E. coli* are positively correlated to a healthy uterine microbiome (Galvao *et al*., 2019; Lima, 2020), which explains their decline in this study. In addition to the abundant uterine pathogens, *Brucella* (*Br.*) *abortus*; *P. anaerobius*; *Peptococcus* and *Peptoniphilus* were also detected from the placental sample. Generally, majority of the taxa detected in the placental sample in this study are often associated with metritis, moreover some abortigenic taxa could also be identified which will be discussed further in subsequent sections.

5.1.3. Analysis of shared and differentially abundant taxa

Analysis of the co-existence and differential abundance of genus-level microbial taxa between the sampled body sites of lactating cows was also conducted in the current study. There were numerous taxa (58 genera) that were commonly shared between the three niches. Of these shared taxa, only three were identical to the 24 shared between three niches in Young *et al*. (2015)'s study *i.e*. *Ruminococcus*, *Turicibacter* and *Coprococcus*. Majority of the shared microbes between the three niches are gut associated. Therefore, the mechanisms involved in the translocation of these gut microbes into milk and blood need to be further investigated. In humans and mice, it has been proven that gut microbes can enter the mammary gland through an endogenous entero-mammary pathway where live bacteria can be transferred from the intestines to the mammary gland via lymphatic and peripheral blood circulation (Rodriguez 2014; Derakhshani *et al*., 2018; Rodriguez *et al*., 2021). Young *et al*., (2015) hypothesized a similar phenomenon by which microbes can be translocated from the gut of cows to the lactating breast via the bloodstream macrophages. However, according to Derakhshani *et al*., (2018) and Rainard, (2017) the link between the immune system of the udder and that of the intestines is very poor in ruminants. They argue that the majority of lymphocytes providing local immunity in

the udder originate from peripheral lymph nodes rather than mucosal sites such as intestines. Furthermore, they state that although hematogenous and lymphatic translocations have been suggested as potential endogenous routes by which certain pathogens *e.g.*, *Mycoplasma* (*M.*) *bovis*, *Mycobacterium* (*My.*) *paratuberculosis* and *Brucella* spp. can cause IMIs and end up being secreted in the milk; it should not be concluded that simultaneous detection of these pathogens in milk and other body sites, particularly body tissues and lymphatic nodes is due to the hypothesized endogenous translocation routes. Rainard, (2017) strongly dispels the existence of the entero-mammary pathway in ruminants but suggests that another possible origin of bacterial DNA in milk other than from breaching of the teat canal from extra-mammary sites could be from the passage of dead bacteria or circulating bacterial components from blood to milk, as large amounts of blood filter through the mammary gland. Nevertheless, as suggested by Rodríguez, (2014) testing of the viability of these microbes through culture and isolation across the body sites is necessary in order to determine if the gut is the source of viable microbial population to the other sites.

Despite 58 genera being shared between the three niches, each genus was represented by multiple bacterial signatures (*i.e.* ASVs) and only a few of these ASVs were found simultaneously across faeces, milk and blood. What could be frequently observed instead, was a pattern of mutual exclusion of ASVs representing the shared taxa between the three niches. That is to say, they appeared to have distinct ecological relationships, with particular clusters of ASVs occurring only in one sample type (*e.g.* faeces) and other clusters in other samples types (*e.g.* milk or blood). In addition to ecological adaptation to the respective host niches, the distinctness of the ASVs representing bacterial taxa across the niches could also be an indication of different species or variants of the same microbe. The observed bacterial signatures that concurrently occurred in corresponding faeces, milk and blood samples from at least one group of animals in a pool included those derived from eight taxa *i.e. Romboutsia*, *UCG-005*, *Prevotellaceae_UCG-004*, *Rikenellaceae_RC9_gut_group*, *Bacteroides*, *Christensenellaceae_R-7_group*, *Turicibacter* and *Fusobacterium*. These bacterial signatures did not match those that were previously reported at OTU level from at least one animal by Young *et al*., (2015) *i.e. Ruminococcus* and *Bifidobacterium* genera as well as an unclassified microbe in the Peptostreptococcaceae family. Owing to the pooling factor in this study, it could not be ascertained that the matching sequence variants across the niches had originated from one animal, but this also does not dispel the possibility of its occurrence. However, the results of this study have the element of biological replication and by pooling, the amount of information that could have been lost below the detection threshold when using individual samples was therefore minimised as previously explained by Schisterman and Vexler, (2008).

Discriminant analysis using DESeq2 showed that there were taxa that were significantly DA between the three niches (*Padj* < 0.01). The phylum Proteobacteria (*Anaplasma*) was significantly enriched in blood than in faeces and milk. Conversely, members of the phylum Firmicutes (*Paeniclostridium, Romboutsia* and *UCG-005*) were greatly reduced in blood than in faeces. Likewise, members of the phylum Verrucomicrobiota (*Akkermansia*), Firmicutes (*Turicibacter*) and Bacteroidota (*Prevotellaceae*_UCG-004) were significantly reduced in blood than in milk.

The most discriminant taxa between the milk and faeces (*i.e*. *Romboutsia*; *Paeniclostridium*; *Monoglobus*; *Akkermansia; Turicibacter; Bacteroides; Candidatus_Saccharimonas*; *UCG-005* and *Prevotellaceae_UCG-004*) were all greatly reduced in milk while significantly enriched in faeces. Members of the phylum Firmicutes, particularly *UCG-005*, *Romboutsia* and *Paeniclostridium* were more significantly enriched than others in faeces.

Among the significantly DA taxa, one genus was greatly enriched in blood *i.e. Anaplasma*. The genus consists of obligate intracellular organisms found exclusively within membrane-bound vacuoles in the cytoplasm of both vertebrate and invertebrate host cells (de La Fuente *et al*., 2005). Bovine anaplasmosis can be attributed mainly to *A. marginale*, *A. centrale* and to a lesser extent *A. bovis* (de La Fuente *et al*., 2005; Khumalo *et al*., 2016; Mutshembele *et al*., 2014). *A. centrale* is less pathogenic and only occasionally associated with clinical disease in cattle than *A. marginale*. As a result, it is presently used as a live vaccine in many countries including South Africa (de La Fuente *et al*., 2005; Khumalo *et al*., 2016).

Anaplasma was the most abundantly detected pathogenic microbe in the entire analysis despite being detected only among the blood group. It was prevalent in 100% of the samples at 73.4% relative abundance. Its very high abundance lead to the initial deduction that what was being detected was the vaccine strain (*A. centrale*). However, upon consultation with the Control Animal Health Technician and the animal owners, there was no history of vaccination of the sampled animals. Without this history, the high abundance per sample was suggestive of an active infection. This prompted further investigation into the obtained blood samples individually in order to characterize the species involved and come up with a better explanation for this stupendous abundance. Furthermore, in order to determine the prevalence of *Anaplasma* among cattle in the sampled study area, blood samples obtained from non-lactating cows present at the dip site were also added to the analysis. The genus was quite abundant among the cattle with a prevalence of 67%. Majority of the sequences corresponded to *A. marginale* (75%) and the remainder to *A. centrale* (25%) field derived strains on the NCBI database. The lack of clinical manifestation of the associated disease among the cattle despite the high infection rate could possibly be attributed to endemic stability. Moreover, infected cattle that survive the acute phase of anaplasmosis remain persistently infected without exhibiting clinical signs despite challenge-

exposure and serve as reservoirs of the pathogen for ticks and other cattle through mechanical transmission of infected blood (Kocan *et al*., 2010).

Recently, a very high abundance of *Anaplasma* has been reported (96.8% of total sequences excluding rare ones) from bovine blood using near full length 16S rRNA sequencing (Kolo *et al*., 2020). In their study conducted on samples collected from Mpumalanga Province, Kolo *et al*., (2020) found that a total of 54% of the sequences were made up of *A. marginale*; followed by different strains of *Anaplasma* spp. that accounted for 40.9% of the abundance; *A. centrale* at 1.4%; *A. platys* at 0.2% and *A. phagocytophilum* at 0.01% abundance. This highlights the significance of characterizing the species of *Anaplasma* from bovine blood as it could be indicative of an active infection and some species within the genus are of zoonotic significance.

The majority of the most discriminant taxa were detected across all three niches but greatly enriched in faeces than in milk and blood. Among these was the genus *Romboutsia* which was detected in 100% of the faecal samples at 26.6% relative abundance. It was also detected in 62.5% of the milk samples at 0.6% relative abundance and in 25% of the blood samples at 0.3% relative abundance. The majority of *Romboutsia*-associated 16S rRNA gene sequences are said to have an intestinal origin in mammals, but the specific roles that they play in the digestive tract remain largely unknown (Gerristen *et al*., 2019). Nonetheless, *Romboutsia* species are said to cover a broad range of metabolic capabilities with respect to carbohydrate utilization, fermentation of single amino acids, anaerobic respiration and metabolic end products (Gerristen *et al*., 2019). The strains differ in their abilities to utilize specific carbohydrates, to synthesize vitamins and other cofactors, and their nitrogen assimilation capabilities (Gerristen *et al*., 2019). *Romboutsia*-like 16S rRNA sequences have been found in intestinal content samples (derived from the duodenum, jejunum, ileum, colon and rectum) of cattle and dogs; in ileal biopsies from humans and pigs and in faecal samples from rats, polar bears, porpoises, humans, rodents and other mammals (Alipour *et al*., 2018; Gerristen *et al*., 2019). In addition, phylotypes of *Romboutsia* have also been detected from teat skin and in high abundances in milk of dairy cattle (Fretin *et al*., 2018).

Interestingly, a single ASV could be resolved to species level among faecal samples in this study and it was a species within the genus *Romboutsia*, *i.e*. *R. sedimentorum* which was first isolated and described from an alkaline-saline lake sediment sample (Wang *et al*., 2015). This ASV was present across all the faecal samples but not in milk and blood although both groups did contain the genus.

Paeniclostridium was also amongst the most abundant taxa and detected in 100% of the faecal samples at 3.6% relative abundance. It was less prevalent in the milk (25% prevalence; 0.3% relative abundance) and in blood (12.5% prevalence; 0.1% relative abundance). Together with *Romboutsia*, they were reportedly the largest genera in heifers of Holstein-Fresian breed and were correlated to digest functions as related to diet composition and probably to physiological

traits (Cendron *et al*., 2020). Likewise, in this study the pair was significantly enriched in faeces. *Paeniclostridium* has also been recently detected in pasteurized milk where it was negatively correlated with the flavour substances, which affected the quality and characteristics of the milk products (Ding *et al*., 2020).

Monoglobus, a newly described genus with a single species in the family Monoglobaceae was also detected across 100% of the faecal samples at a relative abundance of 6.0%. It was also detected in milk (100% prevalence; 2.1% relative abundance) and blood samples (37.5% prevalence; 0.5% relative abundance). The name *Monoglobus pectinilyticus* gen. nov., sp. nov. was proposed by Kim *et al*., in 2017 for the novel Gram-stain-positive, mesophilic, and pectinolytic bacterium isolated from human faeces. *Monoglobus* was unexpected among bovine faecal samples and its detection across the three niches prompted further investigation into its occurrence in the current study. Comparing findings from an earlier version of the taxonomic assignment tool (SILVA v.132) to those obtained in the newer version (SILVA v.138), it was found that top ranking unassigned taxa (at genus level) in the family Ruminococcaceae *i.e. Ruminococcaceae_UCG-010, Ruminococcaceae_UCG-013* and *Ruminococcaceae_UCG-014* were subsequently replaced by the genus *Monoglobus* in the newer version. A throrough search of literature did not yield a record of *Monoglobus* among the microbiota of bovine faeces until recently where Koester *et al*., (2020) tentatively identified a member of the family Oscillospiraceae (*i.e*. *UCG-005*) as being 93.7% similar to *Monoglobus pectinilyticus* in faecal samples of beef cattle. Thus, its role in the sampled niches remains a mystery, in contrast however, members of Ruminococcaceae which it replaces are typically very common and reported to occur among the most abundant taxa found in the bovine GM and in faeces, highlighting their role in digestion of fiber and break down of complex carbohydrates (Chen *et al*., 2017; Fomenky *et al*., 2018).

The genus *Akkermansia* was also significantly enriched in faeces. It was detected at a relative abundance of 6.0% in 100% of the faecal samples. While in milk it was detected in 87.5% of the samples at 2.4% relative abundance and in 25% of the blood samples at 0.4% relative abundance. Dowd *et al*., (2008) detected *Akkermansia* in 95% of cow faecal samples at an abundance ranging between (0.56 - 8.64%), comparable to the findings of the current study with a range of 3.1 - 6.4% across faecal samples. It has been reported in the GM of other ruminants including golden takin at 0.79% relative abundance (Chen *et al.,* 2017); as well as in healthy donkeys where it was the most abundant genus ranging between 17% in females and 23% in males (Liu *et al*., 2014). *Akkermansia* has been found in young calves and thought to play an opportunistic role as the microbe was detected in trace amounts or not detected at all in older animals (Jami *et al*., 2013). The single species genus (*i.e. A. mucinophila*) is said to be an

indicator of healthy breast milk in humans, where it confers beneficial functions as a probiotic (Wu *et al*., 2021). It contributes to a healthy mucus-associated microbial composition and could also be used to prevent obesity and type 2 diabetes (Liu *et al*., 2014). According to a study conducted on mice it is said to have contributed to leanness, moreover, direct supplementation with *Akkermansia* as a probiotic can reportedly result in improvement of glucose intolerance, metabolic endotoxemia, and tissue inflammation (Floch, 2017). Therefore, its abundance in these niches may be important for the health of cows and their offsprings.

Turicibacter was present in 100% of the faecal samples, with an overall abundance of 1.9%. It was also significantly DA and enriched in faeces compared to milk (50% prevalence; 0.4% relative abundance) and blood (12.5% prevalence; <0.1% relative abundance). Similar to the findings of Young *et al.*, (2015), it was also detected across all three niches in this study. The genus has been previously found to be dominant and significantly enriched in many parts of the GIT including the ileum, lumen and large intestine (Mao *et al*., 2015). Cendron *et al*., (2020) reported *Turicibacter* among the most representative genera in faecal samples of heifers and lactating cows. It has also been detected in high abundances in milk of dairy cattle (Fretin *et al*., 2018). The bacterium has been reported to be associated with intestinal butyric acid which stimulates insulin secretion in the pancreas, increases insulin sensitivity, and alters insulin signalling. It has been associated with significant functions such as providing anti-obesity effects, reducing metabolic stress, and inhibiting inflammatory reactions, however, its metabolism and interaction with the host remains unknown (Zhou *et al*., 2019).

Bacteroides was identified in 100% of the faecal and milk samples at 4.0% and 4.1% relative abundances respectively, while detected at 1.8% relative abundance in 75% of the blood samples. *Bacteroides* spp. are well-known intestinal bacteria that can be both beneficial and harmful to their host. Members of this genus play a vital role in the development of immunological tolerance to commensal microbiota and have been noted to participate in natural genetic transfer of antimicrobial resistance genes (Dowd *et al.*, 2008; Malmuthuge *et al*., 2015), which may explain their abundance in new-born calves as opposed to older animals in a study by Jami *et al*., (2013). *Bacteroides* was reported for the first time in cow milk microbiota in 2013 (Quigley *et al.,* 2013), subsequent studies have also characterized it milk (Oikinomou *et al*., 2014; Young *et al*., 2015) including the current. It has also been detected among the blood samples in this study, ranking fifth in terms of abundance.

Overall, various unclassified bacteria (at genus level) derived from Oscillospiraceae (*i.e. UCG-005* at 8.9%); Saccharimonadaceae (*i.e. Candidatus_Saccharimonas* at 2.3%); and Prevotellaceae (*i.e. Prevotellaceae_UCG-004* at 8.4%) families were among the most discriminant taxa and were significantly DA between the three niches. In addition to these, unclassified taxa that were abundantly detected across the three niches although not significantly DA include *Christensenellaceae_R-7_group* at 3.9%; *Prevotellaceae_UCG-001* at 0.6%; *Prevotellaceae_UCG-003* at 2.8%; *Rikenellaceae_RC9_gut_group* at 5.4%; *dgA-11_gut_group* at 2.0% and *Family_XIII_AD3011_group* at 1.0%. Other studies have also previously reported on unclassified fragments from Peptostreptococcaceae, Ruminococcaceae, Enterobacteriaceae, Prevotellaceae, Lachnospiraceae, Oscillospiraceae and Rikenellaceae families as well as from orders Clostridiales and Bacteroidales to predominate the GIT of cattle (Mao *et al*., 2015; Young *et al*., 2015; Koester *et al*. 2020; Huws *et al*., 2011). According to Huws *et al*., (2011), they may play a predominant role in ruminal biohydrogenation. Their frequent reporting and appearance in high abundances signifies the importance of research based on classification of bacterial taxa and updating the database of 16S rRNA gene sequences found in the gut (and ultimately faeces), milk and, blood as well as their roles therein.

5.1.4. Categorization of obtained taxa

Overall, important taxa obtained from this analysis could be summarized under four categories which include: i) arthropod-borne; ii) food-borne and zoonotic; iii) mastitogenic as well as; iv) metritic and abortigenic. The majority of these potentially pathogenic agents fit under more than one category and they were found across all niches, with a few exceptions.

Under the arthropod-borne category *Anaplasma, Bartonella* and *Ehrlichia* were detected in bovine blood. *Anaplasma* has been discussed in detail in earlier sections. *Bartonella*, an emerging arbobacterium genus containing a number of zoonotic pathogens was detected at 25% prevalence and 0.4% relative abundance in this study. The pathogen has also been detected by other authors with the same abundance of 0.4% (Kolo *et al*., 2020) and higher (5.1%) mean relative abundance (Jeon *et al*., 2017) in bovine blood. The genus *Ehrlichia* was detected at a prevalence of 12.5% and 0.2% relative abundance in this study. Recently, Kolo *et al*., (2020) detected *Ehrlichia* species matching a sequence of *E. minasensis* (at ~0.02%) for the first time in South Africa. The most pathogenic and commonly characterized species of *Ehrlichia* infecting wild and domestic ruminants throughout sub-Saharan Africa is *E. ruminantium*, an agent of heartwater (Mtshali *et al*., 2015). It has the potential to spread and cause economic losses if not characterized and properly contained. This calls for implementation of effective control strategies

for ectoparasites such as ticks and fleas which are the possible vectors of the above-mentioned pathogens.

Bacterial genera which consist of pathogenic species associated with the consumption of meat and dairy products were detected across all niches, but were more prevalent in the milk and blood samples in the current study. These include among others *Staphylococcus*, *Clostridium*, *Bacillus, Brucella, Escherichia/Shigella*, *Rhodococccus, Campylobacter*, *Klebsiella* and *Mycobacterium.* Many of these organisms are detected in the GIT of the ruminant animals and can be potential sources of contamination of animal products (Dowd *et al*., 2008; McSweeney and Mackie, 2012; Hoque *et al*., 2019). The chain of events from killing, processing, storage and food preparation is said to provide conducive conditions for multiplication of these contaminating organisms (McSweeney and Mackie, 2012). Some of the species under these genera are generally regarded as secondary contaminants that may contaminate meat during processing *e.g. Clostridium, Staphylococcus* and *Bacillus*. *Clostridium_sensu_stricto_1* is commonly found in cattle manure and the genus contains many medically important species which may have negative effects on animal health depending on the species (Dowd *et al*., 2008). Some species (*e.g. C*. *tetani* and *C. botulinum*) have little ability to invade and multiply in host tissues but produce powerful toxins; while others (*e.g. C. perfringens*, and *C. difficile*) can multiply greatly in tissues and the GIT however produce less potent toxins (Manyi-Loh *et al*., 2016). Likewise, some species of *Staphylococcus* (*e.g. St. aureus*) produce toxins that cannot be destroyed by cooking (McSweeney and Mackie, 2012; Manyi-Loh et *al*., 2016), causing nausea, vomiting and abdominal cramps when infected milk is ingested by humans (Maity and Ambatipudi, 2021). Species of *Bacillus* are mostly harmless and can persist for years in the soil (Manyi-Loh *et al*., 2016), however the most peculiar of the species under the genus is *B. anthracis* that causes anthrax, which according to Manyi-Loh *et al*., (2016) is a life threatening and dreaded disease, especially the pulmonary form which is acquired through inhalation of the bacterial spores. *B. anthracis* and *B. cereus* are currently recognized as pathogens of concern in bovine mastitis due to their increased isolation from milk (Dhanasheker *et al*., 2012; Muehlhoff *et al.*, 2013).

A number of these food-borne pathogens have zoonotic implications, being associated with food poisoning (*e.g. Escherichia, Shigella* spp*.* and *Campylobacter*) and occurrence of human diseases and/or conditions *e.g. Br. abortus*, *M. bovis, My. paratuberculosis* and *Rhodococcus equi* (McSweeney and Mackie, 2012). *Escherichia* (*e.g.* Shiga toxin-producing *E. coli,* O157:H7) and *Shigella* (*e.g. S. flexneri*) are closely related food-borne zoonotic bacteria which cause varying degrees of bacillary dysentery in humans (Dhanasheker *et al*., 2012). In South Africa *S. flexneri* was implicated in several adults and 51% of school children presenting with shigellosis a few hours after eating sour milk contaminated with the pathogen (Dhanasheker *et al*., 2012). *Campylobacter* species (e.g. *Ca. jejuni*) are clinical human and animal pathogens, now

considered to be food-borne pathogens of concern associated with raw milk to which dairy cattle serve as reservoirs (Muehlhoff *et al*., 2013). Members of this genus can cause bovine venereal campylobacteriosis, which is the primary cause of abortion and infertility in cattle (Deng *et al*., 2019). In addition to being classified an agent of contagious abortion and infertility in cattle resulting in significant economic losses, *Br. abortus* causes undulant fever which may progress to a more chronic incapacitating form that can produce serious complications in humans (OIE, 2009). It is obtained through consumption of raw milk and its products as well as through occupational exposure (McSweeney and Mackie, 2012; Nath *et al*., 2013; Frean *et al*., 2018). *Mycobacterium* species (*e.g. My. avium*, *My. paratuberculosis* & *My. bovis*) are responsible for occurrence of bovine and human tuberculosis. These pathogens are transmitted between the environment, wildlife, livestock and humans (commonly through consumption of raw milk) representing major challenges associated with health, economic and sustainable conservation (Biet *et al*., 2005; Dhanasheker *et al*., 2012). On the other hand, *Rhodococcus* (*e.g. R. equi*) acquired through consumption of milk has been implicated in a human mastitis case report (Nath *et al*., 2013). It's also been implicated in cases of clinical (Wani *et al*., 2003) and chronic bovine mastitis (Garg and Kapoor, 1986), but reports from cattle are sparse and mostly describe isolation from purulent lesions in the lymph nodes, particularly in animals with suspected tuberculosis (Witkowski *et al*., 2016). In the current study *Rhodococcus* was also recorded in high abundance in the sample containing the genus *Mycobacterium* than those without, which supports the observation by Witkowski *et al*., (2016).

A number of the identified microbes have been implicated as causative agents of both environmental and contagious mastitis which may culminate in subclinical and clinical forms with serious economic implications for the dairy industry particularly in developing countries like South Africa (Motaung *et al*., 2017). Streptococci (*e.g. Streptococcus uberis* and *S. dysgalactiae*) as well as Gram-negative bacteria such as *Klebsiella* (*e.g. K. oxytoca* and *K. pneumoniae*)*, Escherichia* (*e.g. E. coli*) and *Pseudomonas* (*e.g. P. aeruginosa*) are frequently isolated and have been recognized as the causative agents of environmental bovine mastitis, which is caused by pathogens present in the digestive tract of cows or their surroundings (Ashraf *et al*., 2017; Motaung *et al*., 2017; Pang *et al*., 2018; Hoque *et al*., 2019). In contrast, some species of *Staphylococcus* (*e.g. St. aureus*), *Streptococcus* (*e.g. S. agalactiae*), *Mycoplasma* (*e.g. M. bovis*) and *Corynebacterium* (*e.g. Co*. *bovis*) are recognized as contagious udder pathogens (Ashraf *et al*., 2017; Motaung *et al*., 2017; Hoque *et al*., 2020). These are usually acquired in the milking process through contaminated hands or milking equipment (Motaung *et al*., 2017). Infections with *Klebsiella* spp*., E. coli* and *Streptococcus* spp. cause clinical mastitis, but are often short-lived (Abebe *et al*., 2016). *Pseudomonas* (particularly *P. aeruginosa*) has been previously implicated as a contagious mastitis agent through contaminated milking equipment (Kuehn *et al*., 2013).

However, *Pseudomonas* spp. are more frequently associated with spoilage of dairy products (Kuehn *et al*., 2013; Zhang *et al*., 2015). A number of *Mycoplasma* species have been associated with mastitis, occurring concurrently with arthritis, pneumonia, otitis media and reproductive disorders *e.g.* vulvovaginitis, infertility, endometritis and dystocia (Motaung *et al*., 2017; Parker *et al*., 2018). Although *Corynebacterium* has been previously associated with IMIs, Falentin *et al*., (2016) argue that it may most likely be a colonizer of the teat canal rather than a causal agent of mastitis. Due to the contagious nature of the above-mentioned species it is critical to characterize the species involved in each case. From observation thus far, the GIT harboured and may serve as a reservoir of the food-borne and mastitogenic pathogens, as a number of these were seen in faecal samples, appearing also in blood and milk.

Moreover, metritic and abortigenic taxa were quite prevalent across the three main niches as well as in the placenta. Among these, *Bacteroides* (*e.g. Ba. pyogenes*), *Porphyromonas* (*e.g. Po. levii*), *Prevotella* (*e.g. Pr. melaninogenica*) and *Fusobacterium* (*e.g. F. necrophorum*) are said to account for the majority of infections caused by anaerobic Gram-negative rods (Garret and Onderdonk, 2015). These bacteria particularly have a strong predilection for abscess formation, with the most common sites being the oropharynx; abdominal cavity; lungs; female urogenital tract and also occasionally found in the GIT of both animals and humans (Garret & Onderdonk, 2015; Galvao *et al*., 2017). On the contrary, several species from these genera are useful symbiotic bacteria, facilitating host metabolism and favourably shaping immune responses. However, many of these microbes act opportunistically, causing infections when they gain entry into ‘sterile’ tissues (Garret and Onderdonk, 2015). Among these, although commonly implicated in bovine metritis, *Fusobacterium, Bacteroides* and *Porphyromonas* are said to have a synergistic action with *T. pyogenes* in the aetiology of summer mastitis (Oikonomou *et al*., 2014; Falentin *et al*., 2016). *Staphylococcus* spp. and *Helcococcus* spp. have also been occasionally thrown into the mix (Collins *et al*., 1999). In the current study it was found that milk samples that contained *Fusobacterium* in high abundances also contained high abundances of *Porphyromonas* and the inverse was also true. Likewise, *Trueperella,* although only present in two milk samples. Similar to a previous study (Collins *et al*., 1999), the highest abundance of *Helcococcus* was recorded in samples that also contained *Trueperella* while it was sparsely distributed in those without, which provides further proof of the presence of subclinical mastitis among the sampled cohort. In contrast, *Staphylococcus* and *Bacteroides* were sparsely and randomly distributed without any obvious pattern in relation to the said taxa across the milk samples.

Other than their described synergism in summer mastitis, *F. necrophorum, T. pyogenes, Ba. pyogenes* together with *Pr. melaninogenica* are said to cooperatively cause clinical endometritis in cows and increase the possibility of uterine inflammatory conditions while intensifying disease symptoms (Falentin *et al*., 2016; Vidal *et al*., 2017; Galvao *et al*., 2019; Lima, 2020). Moreover,

Ba. pyogenes is frequently isolated in metritis cases (Lima, 2020); *Fusobacterium* on the other hand has been previously implicated as causing sporadic bovine abortion (Vidal *et al*., 2017); while *Porphyromonas* (*i.e. P. levii*) is of importance in cattle rearing due to its association with bovine necrotic vulvovaginitis and footrot (Dowd *et al*., 2008).

Noteworthy was the detection of genera such as *Actinomyces*, *Anaerococcus*, *Helcococcus*, *Peptococcus*, *Peptostrepotococcus*, *Peptoniphilus* and *Brucella* among the placental microbiota which are known metritic and / or abortigenic taxa. These were mostly unique to the placenta except for *Helcococcus* (also detected in milk), *Peptostreptococcus* (also detected in blood) and *Brucella* (also detected in milk). Galvao *et al*., (2019) listed *Peptoniphilus*, *Peptostreptococcus* and *Helcococcus* among the taxa that are commonly involved in bovine metritis but less prevalent than the above-mentioned frequently isolated genera. *Peptoniphilus* has been previously positively correlated with *Bacteroides, Prevotella, Fusobacterium* and *Porphyromonas* in bovine blood and uterine samples and is also believed to contribute to the development of metritis (Jeon *et al*., 2017). The presence of these microbes in blood shortly after birth in another study was suggestive of the feasibility of a haematogenous spread of uterine pathogens in cows (Jeon *et al*., 2017). Conversely, *Peptostreptococcus* together with *S. pyogenes* and *Fusobacterium* are initial colonizers and are predominant isolates from abscesses (Garret and Onderdonk, 2015). While in addition to their previous detection in subclinical mastitic ovine milk and breast masses (Liu *et al*., 2021), species of *Helcoccocus* have also been isolated from cows that had aborted (*H. ovis*), those with puerperal metritis (*H. ovis* & *H. kunzii*) and valvular endocarditis (*H. ovis*) (Jeon *et al*., 2017; Deng *et al*. 2021); in clinical samples of lower-extremity wounds (*H. kunzii*) and in a human case with pyogenic disease *(H. ovis*), (Deng *et al*. 2021). Galvao *et al*., (2019) proposed that metritis is associated with a dysbiosis of the uterine microbiota characterized by decreased richness, and an increase in Fusobacteriota (particularly *Fusobacterium*) and Bacteroidota (particularly *Bacteroides* and *Porphyromonas*) which suggests that the majority of the cows in this study were metritic. This however could only be confirmed in one cow from which a placental sample was obtained. Ultimately this microbial dysbiosis may have led to the occurrence of the spontaneous expulsion of the fetus. However, after occurrence of an abortion, the placenta is exposed to several environmental contaminants and detection of an agent in the placenta (especially in the case of opportunistic environmental pathogens) does not imply that it was transferred on to the fetus. Thus, even though the placenta as well as blood, milk and faecal samples of this cow show a strong indication of metritic taxa and potential abortifacients (*i.e. Fusobacterium*, *Bacteroides, Porphyromonas*, *Trueperella, Helcococcus* and *Prevotella*), their role in the occurrence of the abortion here remains speculative, nonetheless undisputed.

In addition to the above-mentioned uterine pathogens, *Brucella* was detected in the placental tissue sample. It is the most commonly reported causative agent of spontaneous abortions in cows and due to its perceived zoonotic significance, culturing and species-specific PCR was conducted in the current study in order to characterize the species involved. As stated before, *Brucella* was also detected in two milk samples, one of which was from the same animal with a retained placenta. What could be noted from this analysis was that culturing made it possible to enhance and detect previously undetected pathogens by PCR as could be seen on Fig 4.35 B (lane 2 vs 3), where the same sample had been negative pre-culturing but came out positive post-culturing. Another noteworthy incident was the failure of culturing to enhance *Brucella* from milk samples that constituted a pool that had high microbial counts of this pathogen by 16S rRNA metagenomics. This was possibly because of low bacteraemia. Most importantly was the observation that 16S rRNA metagenomics proved to be more sensitive than both culturing and PCR in detection of *Brucella*, although it could not distinguish between the species involved. However, complemented by culturing, PCR and Sanger sequencing, the species *B. melitensis* and *B. abortus* were detected and characterized in the associated milk sample, while only *B. abortus* could be confirmed from the placental sample.

According to the OIE, (2009), bovine brucellosis is usually caused by *B. abortus*, less frequently by *B. melitensis* and only occasionally by *B. suis* with all three species being highly pathogenic for humans (Ohtsuki *et al*., 2008; Godfroid *et al*., 2010). Following infection with *B. abortus* or *B. melitensis*, pregnant cows develop placentitis usually resulting in abortion between the fifth and ninth month of pregnancy and excretion of the organisms may occur in the milk (OIE, 2009), which explains detection of both species in the milk sample from this study. An abortion and a retained placenta are clinically indicative of bovine brucellosis. Diagnosis depends on the isolation of *Brucella* from abortion material, udder secretions or from tissues removed at post-mortem; moreover, the placenta is considered the most useful sample in determination of the cause of abortion (OIE, 2009; Vidal *et al*., 2017). Although multiple species of *Brucella* were detected from this animal, strictly *B. abortus* could be isolated, amplified and sequenced from the placental tissue thus ruling out the other species and implicating the said species as the cause of abortion in this case. The bacterium could have caused the abortion either singly or in association with the above-mentioned abortifacients which may have acted to intensify the occurrence of the disease. With that being stated however, the impact of mixed infections on abortions remains to be determined (Vidal *et al*., 2017).

Furthermore, looking at the three sample types (faeces, milk and blood) associated with the placental sample in reference to the above-mentioned potential abortifacient and metritic taxa, what could be observed is that the placental microbiota was more similar to the milk microbiota. They contained *Bacteroides*, *Brucella*, *Fusobacterium, Helcococcus*, *Porphyromonas* and

Trueperella in common. *Fusobacterium* was present in all four niches, but *Prevotella* was only detected in the faeces. The faecal sample also contained *Bacteroides*. In contrast *Fusobacterium* and *Bacteroides* were the only two uterine pathogenic genera that could be detected in the associated blood sample. Taking proximity of the urogenital tract and the anus into consideration, it was interesting to observe the distinctness of the faecal microbiota from that of the placenta. It was expected that they would be similar based on findings of a previous study which reported on the similarity of uterine tract microbiota to that of faeces (Jeon *et al*., 2017). The findings of this particular section of the discussion highlight the role of the environment as a potential source of contamination for the teats and the uterine tract and not so much the role of blood in haematogenous transfer of microbes from the gut to the uterine tract (hypothesized by Jeon *et al*., 2017) or the entero-mammary pathway of microbes from the gut to the mammary glands (hypothesized by Young *et al*., 2015).

Lastly, an important observation was made particularly among the milk samples. For the first time ever, this study may have described the milk microbiota of animals with bovine brucellosis which was associated with an increase in metritic pathogens (*i.e. Fusobacterium, Helcococcus, Bacteroides* and *Porphyromonas*) and the presence of unique taxa which although less abundant than the metritic taxa, were only detected in *Brucella* positive samples *i.e. Trueperella*, *Dietzia*, *Facklamia* and an unresolved taxon in the order Bacteroidales. Furthermore, the *Brucella* positive samples, both contained much higher proportions of *Helcococcus, Porphyromonas, Sphingomonas, Narcoidioides, Atopobium, Quadrisphaera* and *Falvobacterium* in comparison to the *Brucella* negative samples. It can be speculated that these taxa are associated with microbial dysbiosis and/or atopobiosis in the milk of animals with bovine brucellosis as the suppression of some and over colonization of other bacterial species in a particular niche results in disease pathogenicity (Deng *et al*., 2019). These findings emphasize the significance of understanding of the interaction between the host's environment and its inhabiting microbes as these microbial interactions may be an important component of disease etiology (Jeon *et al*., 2017; Deng *et al*., 2019).

5.2. CONCLUSION AND RECOMMENDATIONS

DADA2 inference of ASVs was highly resolved, with 98.8% phylum level resolution. The pipeline coupled with SILVA classification database made it possible to identify previously unknown genera in bovine faeces, milk and blood (e.g. *Monoglobus*) and replacing commonly known and abundant unclassified Ruminococcaceae clusters prevalent in the bovine GM.

The tool was efficient for the purpose of the current study, however a significant portion of bacteria could not be assigned at genus-level taxonomy due to unavailability of matches on the sequence database. Therefore, there is room for further research into the taxonomic classification of microbes occurring in faeces, milk and blood. Updating of the database of 16S rRNA gene sequences remains critical and necessary for accurate classification of taxa and subsequent determination of their functions in various body sites.

Furthermore, the sequencing depth achieved in this study proved to be a limiting factor, as a number of sequences generated from the targeted (V3-V4) hypervariable regions were not enough to correctly define some genera and possibly corresponded to reads with unclear sequence accuracy. However, with further improvements and manipulation of the available technologies to their fullest capacity, 16S rRNA metagenomics can yield more sequencing depth and greater species resolution (Jeong *et al*., 2021). For instance, there could be advantage in considering newly developed 16S full-length-based synthetic long-read (sFL16S) method for greater sequencing depth coupled with DADA2 sequence inference to define more bacterial taxa at the genus and species level as it has been found to be more resolute (Jeong *et al*. 2021).

While the targeted hypervariable regions of the 16S rRNA gene or the sequencing depth achieved may not be the optimal for detecting the presence of pathogenic species, the findings of this study indicate that this type of broad scale microbial survey may be useful in determining the presence of potential pathogens from an array of bacteria. This can in turn guide more targeted sampling and detection of both pathogenic and commensal bacteria across body sites.

Due to the complexity and heterogeneity of the available data sets generated by a wide variety of omics platforms, it was difficult to compare and more importantly draw conclusions of the findings of this study based on those of earlier studies. Nevertheless, it was possible to determine the microbial structure of the niches and assess whether their microbial compositions were distinctly homogenous or bore close resemblance to one another through a variety of α- and β-diversity metrics, heatmap cluster, stacked bar plots and UpSetR plot analyses. Furthermore, discriminant analysis served as an important tool to harvest the data generated by sequencing and to identify the bacterial genera that were significantly DA for further analysis in this study.

Microbes can simultaneously occur in the faeces (gut), milk (mammary gland and / or teat canal) and blood (bloodstream) of the same group of animals. The simultaneous occurrence of 16S rRNA bacterial fragments originating from the gut in the blood and milk samples of cows is suggestive of the presence of some endogenous route of transfer of microorganisms from the gut to the mammary glands via the bloodstream of cows as previously hypothesized, however the

findings of this study do not provide definite proof of this. Further investigation into the mechanisms and cells that allow simultaneous detection of microbes in faeces, milk and blood of cows must be conducted. Pooling of the samples may have posed as a limitation to this objective as it could not be ascertained whether an individual animal simultaneously possessed the identified microbes in its faeces, milk and blood, except in the case of animal number 1. Nonetheless this does not take away from the possibility of this occurring and what pooling may have affected is made up for by processing of sample biological replicates simultaneously.

The group of animals used in this study carry/ carried pathogens of veterinary significance and there is a possibility that they are/ were diseased. The abundance and distribution of various types of microbes in different proportions within the sample pools may be associated with particular disease microbial dysbiosis. Furthermore, bacterial genera and species with zoonotic potential were detected. Thus, precaution should be taken to prevent human infection particularly the farmers and herders in the sampled community. Infection can occur orally through ingestion of food products from their cattle, via aerosol due to the proximity of the animal enclosures to their homes and through occupational exposure by handling of infected animals and aborted foetal material which they frequently do (this information on animal husbandry practices was obtained through survey, which was subsequently removed from the analysis).

In conclusion, characterization of the microbiota of the faeces, milk and blood from cattle through high throughput sequencing of the V3-V4 hypervariable region of the 16S rRNA gene provided new insights into the microbiomes of the investigated niches individually and in common, particularly in the South African context.

Now that, to a great extent, the issue of enumeration of microbes which was very limited by classical culture has been overcome, and the very important questions of ‘who is in there and in what proportion?’ have also been answered, going forward the critical step is the determination of ‘what are they doing in there?’. Future studies are thus envisaged to investigate the functionality of the microbiota found in these niches and their potential role in maintaining optimal health and the onset of disease. Furthermore, these studies should be designed with the ‘One-World, One-Health’ approach in mind in order to primarily aid in improvement of productivity through a better understanding of microbial function and ecology. Secondarily, to help to decrease environmental pollution, contamination of food and dissemination of disease between animals and between animals and humans.

BIBLIOGRAPHY

Abebe, R., Hatiya, H., Abera, M., Megersa, B. & Asmare, K. 2016. Bovine mastitis: prevalence, risk factors and isolation of *Staphylococcus aureus* in dairy herds at Hawassa milk shed, South Ethiopia. *BMC Veterinary Research*, 12:270.

Addis, M.F., Tanca, A., Uzzau, S., Oikonomou, G., Bicalho, R.C. & Moroni, P. 2016. The bovine milk microbiota: insights and perspectives from-omics studies. *Molecular BioSystems*, *12*(8), pp.2359-2372.

Alipour, M.J., Jalanka, J., Pessa-Morikawa, T. Kokkonen, T., Satokari, R., Hynönen, U.,…& Niku, M. 2018. The composition of the perinatal intestinal microbiota in cattle. *Scientific Reports*, 8:10437. doi:10.1038/s41598-018-28733-y

Ashraf, A., Imran, M., Yaqub, T., Tayyab, M., Shehzad, W. & Thomson, P.C. 2017. A novel multiplex PCR assay for simultaneous detection of nine clinically significant bacterial pathogens associated with bovine mastitis. *Molecular and cellular probes*, *33*, pp.57-64.

Berrada, Z. L. & Telford, S.R. III. 2009. Burden of tick-borne infections on American companion animals, *Top Companion Animal Medicine*, 24(4):175-181.

Bhatt, V.D., Ahir, V.B., Koringa, P.G., Jakhesara, S.J., Rank, D.N., Nauriyal, D.S., … & Joshi C.G. 2012. Milk microbiome signatures of subclinical mastitis-affected cattle analysed by shotgun sequencing. *Journal of Applied Microbiology*, 112, 639–650.

Biet, F., Boschiroli, M.L., Thorel, M.F. & Guilloteau, L.A. 2005. Zoonotic aspects of Mycobacterium bovis and Mycobacterium avium-intracellulare complex (MAC). *Veterinary Research*, 36(3), pp.411-436.

Bricker B.J., Ewalt D.R., Olsen S.C., Jensen A.E. 2003. Evaluation of the *Brucella abortus* species-specific polymerase chain reaction assay, an improved version of the *Brucella* AMOS polymerase chain reaction assay for cattle. *Journal of Veterinary Diagnostic Investigation*, 15, 374-378.

Bricker, B.J., Halling, S.M. 1994. Differentiation of *Brucella abortus* bv 1, 2 and 4, *Brucella melitensis, Brucella ovis*, and *Brucella suis* bv 1 by PCR. *Journal of Clinical Microbiology*, 32: 2660–6.

Bricker, B.J., Halling, S.M. 1995. Enhancement of the *Brucella* AMOS PCR assay for differentiation of *Brucella* abortus vaccine strains S19 and RB51. *Journal of Clinical Microbiology*, 1995; 33: 1640–42.

Brückner, G.K., Vosloo, W., Du Plessis, B.J.A., Kloeck, P.E.L.G, Connoway, L., Ekron, M.D., ... & Mogajane, M.E. 2002. Foot and mouth disease: the experience of South Africa. *Revue Scientifique et Technique (International Office of Epizootics)*, 21 (3):751-764.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7):581-3. Epub 2016/05/23. doi: 10.1038/nmeth.3869. PubMed PMID: 27214047

Carrillo-Casas, E.M. & Miranda-Morales, R. E. 2012. *Bovine Mastitis Pathogens: Prevalence and Effects on Somatic Cell Count In: Milk Production – An Up-to-Date Overview of Animal Nutrition, Management and Health*. Intech Open Science, http://dx.doi.org/10.5772/51032

Castillo, D.J., Rifkin, R.F., Cowan, D.A. & Potgieter, M. 2019. The healthy human blood microbiome: Fact or Fiction? *Frontiers in Cellular Infection and Microbiology*, 9:148. doi: 10.3389/fcimb.2019.00148

Cendron, F., Niero, G., Carlino, G., Penasa, M. & Cassandro, M. 2020. Characterizing the fecal bacteria and archaea community of heifers and lactating cows through 16S rRNA next-generation sequencing. *Journal of Applied Genetics*, 61(4), pp.593-605.

Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods*, **69**(2): p. 330-9.

Chapman, M., Warren, E.H., III. & Wu, C.J. 2012. Applications of Next Generation Sequencing to blood and marrow transplantation. *Biology of Blood and Marrow Transplantation*, 18(1 Suppl): S151–S160. doi:10.1016/j.bbmt.2011.11.011

Chen, J., Zhang, H., Wu, X., Shang, S., Yan, J., Chen, Y., …& Tang, X. 2017. Characterization of the gut microbiota in the golden takin (*Budorcas taxicolor bedfordi*). *AMB Express*, *7*(1), pp.1-10.

Chen, K. & Pachter, L. 2005. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLOS Computational Biology*, 1(2): e24.

Clarridge, J.E. III. 2004. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews*, 17(4 0893), 840–862.

Collins, M.D., Falsen, E., Foster, G., Monasterio, L.R., Dominguez, L. & Fernandez-Garazabal, J.F. 1999. *Helcococcus ovis* sp. nov., a gram-positive organism from sheep. *International Journal of Systematic Bacteriology,* 4:1429-1432.

Conway, J.R., Lex, A. & Gehlenborg, N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33(18):2938-40. doi: 10.1093/bioinformatics/btx364

Crowder, C.D., Rounds, M.A., Phillipson C. A., Picuri, M.J., Matthews, H. E., Halverson, J, … & Eshoo M.W. 2010. Extraction of total nucleic acids from ticks for the detection of bacterial and viral pathogens. *Journal of Medical Entomology*, 47(1): 89–94.

DAFF (Department of Agriculture Forestry and Fisheries). 2012. *A profile of the South African dairy market value chain*. https://www.nda.agric.za/docs/AMCP/Dairy2012.pdf Date of access: 03 May. 2020.

DAFF (Department of Agriculture Forestry and Fisheries). 2018. *A profile of the South African beef market value chain*. https://www.nda.agric.za/docs/AMCP/Beef2012-13.pdf Date of access: 03 May. 2020.

De La Fuente, J., Massung, R.F., Wong, S J., Chu, F.K., Lutz, H., Meli, M., … &, Kocan, K.M. 2005. Sequence analysis of the msp4 gene of *Anaplasma phagocytophilum* strains. *Journal of Clinical Microbiology*, 43, 1309–1317.

De Leoz, M.L.A., Kalanetra, K.M. Bokulich, N.A., Strum, J.S. Underwood M.A., German J.B., … & Lebrilla, C.B. 2015. Human milk glycomics and gut microbial genomics in infant feces show a correlation between human milk oligosaccharides and gut microbiota: A proof-of-concept study. *Journal of Proteome Research*, 14, 491−502. dx.doi.org/10.1021/pr500759e

Deng, F., McClure, M., Rorie, R., Wang, X., Chai, J., Wei, X., … & Zhao, J. 2019. The vaginal and fecal microbiomes are related to pregnancy status in beef heifers. *Journal of Animal Science and Biotechnology*, 10 (1), pp.1-13.

De Oliveira, G.B.D., Favarin, L., Luchese, R.H. & McIntosh, D., 2015. Psychrotrophic bacteria in milk: How much do we really know?. *Brazilian Journal of Microbiology*, 46, pp.313-321.

Derakhshani, H., Plaizier, J. C., De Buck, J., Barkema, H. W. & Khafipour, E. 2018. Composition of the teat canal and intramammary microbiota of dairy cows subjected to antimicrobial dry cow therapy and internal teat sealant. *Journal of Dairy Science*, 101, 10191–10205. doi:10.3168/jds.2018-14858

Dhanashekar, R., Akkinepalli, S. & Nellutla, A. 2012. Milk-borne infections. An analysis of their potential effect on the milk industry. *GERMS*, 2(3):101-109

Dias A.L.G. & Ametaj B.N. 2017. *Cattle Gastrointestinal Tract Microbiota in Health and Disease*. In: Ametaj B. (eds) Periparturient Diseases of Dairy Cows. Springer, Cham. https://doi.org/10.1007/978-3-319-43033-1_5

Ding, R., Liu, Y., Yang, S., Liu, Y., Shi, H., Yue, X., … & Wu, J. 2020. High-throughput sequencing provides new insights into the roles and implications of core microbiota present in pasteurized milk. *Food Research International*, *137*, p.109586.

Dowd, S.E., Callaway, T.R., Wolcott, R.D., Sun, Y., McKeehan, T., Hagevoort, R.G. & Edrington, T.S. 2008. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiology*, 8(1), pp.1-8.

Ducrotoy, M, Bertu, W.J., Matope, G., Cadmus, S., Conde-Alverez, R., Gusi, A.M., … & Moriyon, I. 2015. Brucellosis in Sub-Saharan Africa: Current challenges for management, diagnosis, and control. *Acta Tropica* 165 2017: 179-193.

Epstein J.H. & Price J.T. 2009. The significant but understudied impact of pathogen transmission from humans to animals. *Mount Sinai Journal of Medicine* 76:448–455.

Eisenhofer, R., Minich, J.J., Marotz, C., Cooper, A., Knight, R., Weyrich, L.S. 2019. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends in Microbiology*, 27(2):105-17.

Falentin, H., Rault, L., Nicolas, A., Bouchard, D.S., Lassalas, J., Lamberton, … & Even, S. 2016. Bovine teat microbiome analysis revealed reduced alpha diversity and significant changes in taxonomic profiles in quarters with a history of mastitis. *Frontiers in Microbiology*, 7:480. doi: 10.3389/fmicb.2016.00480

Faria, M.M.P, Conly, J.M., Surette, M.G. 2015. The development and application of a molecular community profiling strategy to identify polymicrobial bacterial DNA in the whole blood of septic patients. *BMC Microbiology*, 15:215 doi10.1186/s12866-015-0557-7

Farrell, M. J., Govender, D., Hajibabaei, M., Van Der Bank, M., & Davies, T. J. 2019. Bacterial diversity in the waterholes of the Kruger National Park: an eDNA metabarcoding approach. *Genome*, 62(3), 229-242.

Floch, N. 2017. The Influence of Microbiota on Mechanisms of Bariatric Surgery. *The Microbiota in Gastrointestinal Pathophysiology* (pp. 267-281). Academic Press.

Fomenky, B.E., Do, D.N., Talbot, G., Chiquette, J., Bissonnette, N., Chouinard, Y.P., ... & Ibeagha-Awemu, E.M. 2018. Direct-fed microbial supplementation influences the bacteria community composition of the gastrointestinal tract of pre-and post-weaned calves. *Scientific Reports*, *8*(1), pp.1-21.

Food and Agriculture Organization – United Nations (FAO). 2014. Impact of mastitis in small scale dairy production systems. Animal Production and Health Working Paper. No.13. Rome. http://www.fao.org/3/i3377e/i3377e.pdf Date of access 18 June. 2019.

Franco-Lopez, J., Duplessis, M., Bui, A., Reymond, C., Poisson, W., Blais, L., Chong, J., Gervais, R., Rico, D.E., Cue, R.I., Girard, C.L., Ronholm, J. 2020. Correlations between the composition of the bovine microbiota and vitamin B12 abundance. *Systems*, 5:e00107-20. https://doi.org/10.1128/mSystems.00107-20

Frean, J., Cloete, A., Rossouw, J. & Blumberg, L. 2018. Brucellosis in South Africa - A notifiable medical condition. *NICD Communicable Diseases Communique*, 16(3), pp.110-117.

Frétin, M., Martin, B., Rifa, E., Isabelle, V.M., Pomiès, D., Ferlay, A., … & Delbès, C., 2018. Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems. *Scientific Reports*, 8(1), pp.1-11.

Galimanas, V., Hall, M.W., Singh, N., Lynch, M.D.J., Goldberg, M., Tenenbaum, …& Senadheera, D.B. 2014. Bacterial community composition of chronic periodontitis and novel oral sampling sites for detecting disease indicators. *Microbiome*, *2*(1), pp.1-13.

Galvão, K.N., Bicalho, R.C. & Jeon, S.J. 2019. Symposium review: the uterine microbiome associated with the development of uterine disease in dairy cows. *Journal of Dairy Science*, 102(12), pp.11786-11797.

Ganda, E.K., Bisinotto, R.S., Lima, S.F., Kronauer, K., Decter, D.H., Oikonomou, G., … Bicalho, R.C. 2016. Longitudinal metagenomics profiling of bovine milk to assess the impact of intramammary treatment using a third-generation cephalosporin. Antimicrobial usage in food animals has a direct impact on human health. *Scientific Reports*, 6:37565.

Garg, D.N., & Kapoor, P.K. 1986. Isolation and characterization of *Rhodococcus* (*Corynebacterium*) *equi* from cows with mastitis. *Indian Journal of Comparative Microbiology, Immunology and Infectious Diseases*, 7 (2/3), pp. 91-95.

Garrett, W.S. & Onderdonk, A.B. 2015. *Bacteroides, Prevotella, Porphyromonas*, and *Fusobacterium* species (and other medically important anaerobic Gram-negative bacilli), In *Mandell, douglas, and bennett's principles and practice of infectious diseases, 8th* Ed, Editor(s): Bennett, J.E., Dolin R. & Blaser, M.J., W.B. Saunders, Pages 2773-2780, ISBN 9781455748013, https://doi.org/10.1016/B978-1-4557-4801-3.00249-6

Gerritsen, J., Hornung, B., Ritari, J., Paulin, L., Rijkers, G.T., Schaap, P.J., … & Smidt, H. 2019. A comparative and functional genomics analysis of the genus *Romboutsia* provides insight into adaptation to an intestinal lifestyle. *BioRxiv*, p.845511.

Gibbons, L.M., Jacobs, E.M., Hansen, J. 2020. Faecal examination of farm animals for helminth parasites. Royal Veterinary College/Food Agricultural Organization Guide to Veterinary Diagnostic Parasitology https://www.rvc.ac.uk/Review/Parasitology/Faeces/Step1a.htm#

Grada, A. & Weinbrecht, K. 2013. Next-Generation Sequencing: Methodology and Application. *Journal of Investigative Dermatology,* 133, e11. doi:10.1038/jid.2013.248

Godfroid, J., Nielsen, K. & Saegerman, C. 2010. Diagnosis of brucellosis in livestock and wildlife. *Croatian medical journal*, 51(4), pp. 296-305.

Guo, H., Moumounia, P.F.A., Thekisoe, O., Gaoa, Y., Liua, M.,Lia, ... & Xuana, X. 2019. Genetic characterization of tick-borne pathogens in ticks infesting cattle and sheep from three South African provinces. *Ticks and Tick-borne Diseases*. doi: 10.1016/j.ttbdis.2019.04.008

Gupta, N., Verma, V.K. 2019. Next-Generation Sequencing and its application: Empowering in public health beyond reality. In: Arora P. (eds) *Microbial technology for the welfare of society. Microorganisms for Sustainability*, vol 17. Springer Singapore, pp 313-341. https://doi.org/10.1007/978-981-13-8844-6_15

Gupta, S., Mortensen, M.S., Schjørring, S., Trivedi, U., Vestergaard, G., Stokholm, J., … et al. 2019. Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Communications biology*, 2(1):1-7.

Kuczynski, J., Stombaugh, J., Walters, W.A., González, A., Caporaso, J.G., Knight R. 2011. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Current Protocols in Bioinformatics*, https://doi:10.1002/0471250953.bi1007s36

Halajian, A., Palomar, A.M., Portillo, A., Heyne, H., Luus-Powell, W.J. & Oteo, J.A. 2016. Investigation of *Rickettsia, Coxiella burnetii* and *Bartonella* in ticks from animals in South Africa', *Ticks and Tick-borne Diseases*, 7, 361–366. https://doi.org/10.1016/j.ttbdis.2015.12.008

Higuchi, B.S., Rodrigues, N., Gonzaga, M.I., Paiolo, J.C.C., Stefanutto, N., Omori, W.P., …& De Oliveira, G.L.V. 2018. Intestinal dysbiosis in autoimmune diabetes is correlated with poor glycemic control and increased interleukin-6: a pilot study. *Frontiers in immunology*, 1689.

Hodkinson, B.P. & Grice E.A. 2014. Next-Generation Sequencing: a review of technologies and tools for wound microbiome research. *Advances In Wound Care*, (4):1. doi: 10.1089/wound..0542

Hoque, M.N., Istiaq, A., Clement, R.A., Sultana, M., Crandall, K.A., Siddiki, A.Z. & Hossain, M.A. 2019. Metagenomic deep sequencing reveals association of microbiome signature with functional biases in bovine mastitis. *Scientific Reports*, 9:13536. https://doi.org/10.1038/s41598-019-49468-4

Huws, S.A., Kim, E.J., Lee, M.R., Scott, M.B., Tweed, J.K., Pinloche, E., … & Scollan, N.D., 2011. As yet uncultured bacteria phylogenetically classified as *Prevotella*, Lachnospiraceae incertae sedis and unclassified Bacteroidales, Clostridiales and Ruminococcaceae may play a predominant role in ruminal biohydrogenation. *Environmental Microbiology*, 13(6), pp.1500-1512.

Jami, E., Israel, A., Kotser, A. & Mizrahi, I. 2013. Exploring the bovine rumen bacterial community from birth to adulthood. *The International Society for Microbial Ecology Journal*, 7(6), pp.1069-1079.

Jami, E., White, B.A. & Mizrahi, I. 2014. Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *PLoS ONE*, 9(1): e85423. doi:10.1371/journal.pone.0085423

Janda, J.M. & Abbott, S.L., 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, 45(9), pp.2761-2764.

Jeon, S.J., Cunha, F., Vieira-Neto, A., Bicalho, R.C., Lima, S., Bicalho, M.L. & Galvão, K.N. 2017. Blood as a route of transmission of uterine pathogens from the gut to the uterus in cows. *Microbiome*, 5 (1): 109.

Jeong, J., Yun, K., Mun, S., Chung, W.H., Choi, S.Y., Lim, M.Y., …& Han, K. 2021. The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology. *Scientific reports*, *11*(1), pp.1-12.

Jongejan, F. & Uilenberg, G. 2004. The global importance of ticks. *Parasitology,* 129: S3–S14.

Kaoud H. A. 2019. *Introductory Chapter: Bacterial Cattle Diseases - Economic Impact and Their Control.* http://dx.doi.org/10.5772/intechopen.83635

Khasapane, N.G., Nkhebenyane, S.J., Thekisoe, O. Taioe, M. O., Kwenda, S., Mtshali, P.S. & Khumalo, Z.T.H. 2020. Characterisation of raw milk microbiome from Harrismith, Free-State Using 16S rRNA Gene Sequencing. *Proceedings of the IRES International Conference*, Johannesburg, South Africa, 25th – 26th March 2020.

Khumalo, Z.T., Catanese, H.N., Liesching, N., Hove, P., Collins, N.E., Chaisi, M.E., … & Brayton, K.A. 2016. Characterization *of Anaplasma marginale* subsp. *centrale* strains by use of msp1aS genotyping reveals a wildlife reservoir. *Journal of Clinical Microbiology*, 54(10), pp.2503-2512.

Kirkan, Ş, Erbaş G. & Parin U. 2017. *Chapter 7: Bacterial Tick-Borne Diseases of Livestock Animals*. IntechOpen, http://dx.doi.org/10.5772/66122. Date of access: 25 June. 2020.

Kirsten J.F. & Van Zyl J. 1998. Defining small-scale farmers in the South African context. *Agrekon*, 37(4).

Klimesova, K., Jiraskova Zakostelska, Z &Tlaskalova-Hogenova, H. 2018. Oral bacterial and fungal microbiome impacts colorectal carcinogenesis. *Frontiers in Microbiology,* 9:774. doi: 10.3389/fmicb.2018.00774

Klindworth, A., Pruesse, E., Schweer,T., Peplies, J., Quast, C., Horn, M. & Glockner, F.O. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*. 41 (1) e1 doi:10.1093/nar/gks808

Koester, L.R., Poole, D.H., Serão, N.V.L., Schmitz-Esser, S. 2020. Beef cattle that respond differently to fescue toxicosis have distinct gastrointestinal tract microbiota. *PLoS One*, 15(7): e0229192. https://doi.org/10.1371/journal. pone.0229192

Kolbe, A.R., Castro-Nallar, E., Preciado, D. & Pérez-Losada, M. 2019. Altered middle ear microbiome in children with chronic otitis media with effusion and respiratory illnesses. *Frontiers in Cellular and Infection Microbiology*, 9, p.339.

Kolo, A.O., Collins, N.E., Brayton, K.A., Chaisi, M., Blumberg, L., Frean, J., ….& Weyer, J., 2020. *Anaplasma phagocytophilum* and other *Anaplasma* spp. in various hosts in the Mnisi Community, Mpumalanga Province, South Africa. *Microorganisms*, 8(11), p.1812.

Kuczynski, J., Stombaugh, J., Walters, W.A., González, A., Caporaso, J.G., Knight R. 2011. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Current Protocols in Bioinformatics*. doi:10.1002/0471250953.bi1007s36

Kuehn, J.S., Gorden P.J., Munro D., Rong R., Dong, Q., Plummer, P.J., …& Phillips, G.J. 2013. Bacterial community profiling of milk samples as a means to understand culture-negative bovine clinical mastitis. *PLoS ONE*, 8(4): e61959.

Kwong, J.C., Mccallum, N., Sintchenko, V., & Howden, B. P. 2015. Whole genome sequencing in clinical and public health microbiology. *Pathology*, 47(3), pp. 199–210.

Ledwaba, M.B., Ndumnego, O.C., Matle, I., Gelaw, A.K. & Van Heerden, H. 2020. 'Investigating selective media for optimal isolation of *Brucella* spp. in South Africa', *Onderstepoort Journal of Veterinary Research*, 87(1), a1792. https://doi.org/ 10.4102/ojvr.v87i1.1792

Lima, F.S. 2020. Recent advances and future directions for uterine diseases diagnosis, pathogenesis, and management in dairy cows. *Animal Reproduction,* 17(3):e20200063. https://doi.org/10.1590/1984-3143-AR2020-0063

Lima, S.F., Teixeira, A.G.V., Lima, F.S., Ganda, E.K., Higgins, C.H., Oikonomou, G., & Bicalho R.C. 2017. The bovine colostrum microbiome and its association with clinical mastitis. *Journal of Dairy Science,* 100 (4) :3031–3042. https://doi.org/10.3168/jds.2016-11604

Lindner, J.D.D., Santarelli, M., Yamaguishi, C.T., Soccol, C.R. & Neviani E. 2011. Recovery and identification of bovine colostrum microflora using traditional and molecular approaches: bovine colostrum microflora recovery. *Food Technology and Biotechnology*, 49 (3) 364–368.

Liu, K., Deng, Z., Zhang, L., Gu, X., Liu, G., Liu, Y., … & Qu, W. 2021. The first report and biological characteristics of *Helcococcus ovis* isolated from clinical bovine mastitis in a Chinese dairy herd. *bioRxiv*.

Liu, X., Fan, H., Ding, X., Hong, Z., Nei, Y., Liu, Z., … & Guo, H. 2014. Analysis of the gut microbiota by high-throughput sequencing of the V5–V6 regions of the 16S rRNA gene in donkey. *Current Microbiology*, 68(5), pp.657-662.

Loohuis, L.M.O., Mangul, S., Ori, A. P., Jospin, G., Koslicki, D., Yang, H.T., ... & Ophoff, R.A. 2018. Transcriptome analysis in whole blood reveals increased microbial diversity in schizophrenia. *Translational Psychiatry,* 8:96. doi: 10.1038/s41398-018-0107-9

Lorusso, V., Wijnveld, M., Majekodunmi, A.O., Dongkum, C., Fajinmi, A. Dogo, A.G., …& Picozzi, K. 2016.Tick-borne pathogens of zoonotic and veterinary importance in Nigerian cattle. *Parasites & Vectors*, 9:217. doi:10.1186/s13071-016-1504-7

Love, M.I., Huber, W., Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550. doi: 10.1186/s13059-014-0550-8

Maity, S. & Ambatipudi, K. 2021. Mammary microbial dysbiosis leads to the zoonosis of bovine mastitis: a One-Health perspective. *FEMS Microbiology Ecology*, 97(1), p.fiaa241.

Malmuthuge, N., Griebel, P.J. & Guan, L.L. 2015. The gut microbiome and its potential role in the development and function of newborn calf gastrointestinal tract. *Frontiers in Veterinary Science*, 2:36. doi: 10.3389/fvets.2015.00036

Mandal, R.K., Jian, T., Al-Rubaye, A.A., Rhoads, D.D., Wideman, R.F., Zhao, J., ... & Young, M.K. 2016. An investigation into blood microbiota and its potential association with Bacterial Chondronecrosis with Osteomyelitis (BCO) in broilers. *Scientific Reports*, 6:25882. doi: 10.1038/srep25882

Manyi-Loh, C.E., Mamphweli, S.N., Meyer, E.L., Makaka, G. Simon, M., and Okoh, A.I. 2016. An overview of the control of bacterial pathogens in cattle manure. *International Journal of Environmental Research and Public Health*, 13, 843.

Mao, S., Zhang, M., Liu, J., & Zhu, W. 2015. Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. *Scientific Reports*, *5*, 16116.

Marufatuzzahan, P.S, Chowdhury, T.A., Shanzana, P., Moli, M.A., Zahan, F.N. & Rafa, R.T. 2018. A questionnaire survey on common animal husbandry and hygiene practices among the small scale livestock farmers in suburban area of Sylhet, Bangladesh. *Agricultural & Veterinary Sciences*, 2(1), pp.38-48 38.

Matjila, P.T., Leisewitz, A.L., Jongejan, F. & Penzhorn, B.L. 2008. Molecular detection of tick-borne protozoal and ehrlichial infections in domestic dogs in South Africa. *Veterinary Parasitology*, 155: 152–157.

McDaniel, C.J., Cardwell, D.M., Moeller, R.B. & Gray, G.C. 2014. Humans and Cattle: A Review of Bovine Zoonoses. *Vector-Borne and Zoonotic Diseases*,14:(1).

McDermott, J.J. & Arimi. S.M. 2002. Brucellosis in sub-Saharan Africa: epidemiology, control and impact. *Veterinary Microbiology*, 90(1-4), pp.111-134.

McLaren, M.R. 2020. Silva SSU taxonomic training data formatted for DADA2 (Silva version 138) (Version 2) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3986799

McMurdie & P.J. &Holmes, S. 2013. Phyloseq: An R Package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8(4): e61217. doi:10.1371/journal.pone.0061217

McMurdie, P.J. & Holmes, S. 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology*, 10(4): e1003531. doi:10.1371/ journal.pcbi.1003531

McSweeney, C. & Mackie, R. 2012. *Micro-organisms and ruminant digestion: state of knowledge, trends and future prospects.* Commission on Genetic Resources for Food and Agriculture. Background Study Paper *(FAO)* No. 6. 61, pp.1-62.http://www.fao.org/3/me992e/me992e.pdf Date of access: 29 September. 2020.

Milanese, A., Mende, D.R., Paoli, L., Salazar, G., Ruscheweyh, H-J., Cuenca, M., ... & Sunagawa, S. 2019. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nature Communications*, 10:1014. doi.org/10.1038/s41467-019-08844-4

Moon, C.D., Young, W., Maclean, P.H., Cookson, A.L. & Bermingham, E.N. 2018. Metagenomic insights into the roles of Proteobacteria in the gastrointestinal microbiomes of healthy dogs and cats. *Microbiology Open*, *7*(5), p.e00677.

Motaung, T.E., Petrovski, K.R., Petzer, I-M., Thekisoe, O., Tsilo, T.J. 2017. Importance of bovine mastitis in Africa. *Animal Health Research Reviews*, Cambridge University Press. doi:10.1017/S1466252317000032

Mtshali, K., Khumalo, Z.T., Nakao, R., Grab, D.J., Sugimoto, C. & Thekisoe, O.M., 2015. Molecular detection of zoonotic tick-borne pathogens from ticks collected from ruminants in four South African provinces. *Journal of Veterinary Medical Science*, pp.15-0170.

Muehlhoff, E., Bennett, A. & McMahon, D. 2013. *Milk and Dairy Products in Human Nutrition*. Food and Agriculture Organization of the United Nations (FAO), Rome.

Nath, S.R., Mathew, A.P., Mohan, A. & Anila, K. R. 2013. *Rhodococcus equi* granulomatous mastitis in an immunocompetent patient, *Journal of Medical Microbiology* 62, 1253–1255, doi 10.1099/jmm.0.054346-0

Ndhlovu, D.N., Makaya, P.V. & Penzhorn, B.L. 2009. Tick infestation, and udder and teat damage in selected cattle herds of Matabeleland South, Zimbabwe, *Onderstepoort Journal Veterinary Research,* 76:235–248.

O'Hara, E., Neves, A.L.A, Song, Y. & Guan, L.L. 2020. The role of the gut microbiome in cattle production and health: Driver or passenger? *Annual Review of Animal Biosciences*, 8:199–220.

Ohtsuki, R., Kawamoto, K., Kato, Y., Shah, M.M., Ezaki, T. & Makino, S.I. 2008. Rapid detection of *Brucella* spp. by the loop-mediated isothermal amplification method. Journal of Applied Microbiology, 104(6), pp.1815-1823.

Oikonomou, G., Addis, M.F. Chassard, C., Nader-Macias, M.E.F.N., Grant, I., Delbès, C., … & Even, S. 2020. Milk microbiota: what are we exactly talking about? *Frontiers in Microbiology*, 11:60. doi: 10.3389/fmicb.2020.00060

Oikonomou, G., Bicalho, M. L., Meira, E., Rossi, R. E., Foditsch, C., Machado, V. S., … & Bicalho, R.C. 2014. Microbiota of cow's milk; distinguishing healthy, sub-clinically and clinically diseased quarters. *PLoS One*, 9:e85904. doi: 10.1371/journal.pone. 0085904

Oikonomou, G., Machado, V.S., Santisteban, C., Schukken, Y.H., & Bicalho, R.C. 2012. Microbial diversity of bovine mastitic milk as described by pyrosequencing of metagenomic 16s rDNA. *PLoS One,* 7:e47671. doi: 10.1371/ journal.pone.0047671

Oikonomou, G., Teixeira, A.G.V., Foditsch, C., Bicalho, M.L., Machado, V.S. & Bicalho, R.C. 2013. Fecal microbial diversity in pre-weaned dairy calves as described by pyrosequencing of metagenomic 16S rDNA. Associations of *Faecalibacterium* species with health and growth. *PloS One*, 8(4), p.e63157.

Ondov, B.D., Bergman, N.H., Phillippy, A.M. 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*. 12:385. Epub 2011/10/04. doi: 10.1186/1471-2105-12-385. PubMed PMID: 21961884; PubMed Central PMCID: PMCPMC3190407

Ong, C.T., Turni, C., Blackall, P.J., Boe-Hansen, G., Hayes, B.J. & Tabor, A.E. 2021. Interrogating the bovine reproductive tract metagenomes using culture-independent approaches: a systematic review. *Animal Microbiome*, 3(1), pp.1-12.

Oyama, L.B., Girdwood, S.E. Cookson, A.R., Fernandez-Fuentes, N., Privé F., Vallin, H.E., …& Huws, S.A. 2017. The rumen microbiome: an underexplored resource for novel antimicrobial discovery. *Biofilms and Microbiomes*, 3:33. doi:10.1038/s41522-017-0042-1

Panek, M., Paljetak, H.C., Barešić, A., Perić, M., Matijašić, M., Lojkić, I., … & Verbanac, D. 2018. Methodology challenges in studying human gut microbiota - effects of collection, storage, DNA extraction and next generation sequencing technologies. *Scientific Reports,* 8:5143. doi:10.1038/s41598-018-23296-4

Pang, M., Xie, X., Bao, H., Sun, L., He, T., Zhao, H., … & Xie, K. 2018. Insights into the bovine milk microbiota in dairy farms with different incidence rates of subclinical mastitis. *Frontiers in Microbiology*, 9, p.2379.

Parker, A.M., Sheehy, P.A., Hazelton, M.S., Bosward, K.L. & House, J.K. 2018. A review of *Mycoplasma* diagnostics in cattle. *Journal of Veterinary Internal Medicine*, 32(3), pp.1241-1252.

Pearce, M.C., Fenlon, D., Low, J.C., Smith, A.W., Knight, H.I., Evans, J., … & Gunn, G.J. 2004. Distribution of *Escherichia coli* O157 in bovine fecal pats and its impact on estimates of the prevalence of fecal shedding. *Appl Environ Microbiol,* 70, 5737–5743.

Perez, P.F., Dore, J., Leclerc, M., Levenez, F., Benyacoub, J., Serrant, P., … & Donnet-Hughes A. 2007. Bacterial imprinting of the neonatal immune system: lessons from maternal cells? *Pediatrics*, 119:e724–e732. doi: 10.1542/peds.2006-1649

Potgieter, M., Bester, J., Kell, D.B. and Pretorius, E., 2015. The dormant blood microbiome in chronic, inflammatory diseases. *FEMS microbiology reviews*, 39(4), pp.567-591.

Porter, J.W.G. 1969. Digestion in the pre-ruminant animal. *Proceedings of the Nutrition Society*, *28*(1), pp.115-121.

Quigley, L., McCarthy, R., O'Sullivan, O., Beresford, T.P. Fitzgerald G.F., Ross, R. P., Stanton, C. & Cotter, P.D. 2013. The microbial content of raw and pasteurized cow milk as determined by molecular approaches. *Journal of Dairy Science,* 96:4928–4937.

Rainard, P. 2017. Mammary microbiota of dairy ruminants: fact or fiction? *Veterinary Research*, 48(1), pp.1-10.

Rampelli, S., Candela, M., Turroni, S., Biagi, E., Collino, S., Franceschi, C., … & Brigidi, P. 2013. Functional metagenomic profiling of intestinal microbiome in extreme ageing. *AGING*, 5(12).

Rapp, D. 2010. DNA extraction from bovine faeces: current status and future trends. The Society for Applied Microbiology. *Journal of Applied Microbiology,* 108:1485–1493.

Rausch, P., Rühlemann, M., Hermes, B.M.; Doms, S.; Dagan, T., Dierking, K., …& Baines, J.F. 2019. Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome*, 7:133. doi:10.1186/s40168-019-0743-1

Reuter, J.A., Spacek, D.V. & Snyder, M.P. 2015. High-throughput sequencing technologies. *Molecular Cell, Elsevier Inc*. 58. http://dx.doi.org/10.1016/j.molcel.2015.05.004

Robinson, S. E., Brown, P. E., Wright, E. J., Hart, C. A. & French, N. P. 2009. Quantifying within- and between-animal variation and uncertainty associated with counts of *Escherichia coli* O157 occurring in naturally infected cattle faeces. *Journal of The Royal Society Interface*, 6, 169–177. doi:10.1098/rsif.2008.0183

Robinson, S.E., Brown, P.E., Wright, E.J., Bennett, M., Hart, C.A. & French, N.P. 2005. Heterogeneous distributions of *Escherichia coli* O157 within naturally infected bovine faecal pats. *FEMS Microbiology Letters,* 244, 291–296.

Rodríguez J.M. 2014. The origin of human milk bacteria: is there a bacterial entero-mammary pathway during late pregnancy and lactation?. *Advances in Nutrition*, 5(6):779-84.

Rodríguez, J.M., Fernández, L. & Verhasselt, V. 2021. The Gut–Breast Axis: Programming Health for Life. *Nutrients*, 13(2), p.606.

Ross, E.M., Moate, P.J, Bath, C.R., Davidson, S.E., Sawbridge, T.I., Guthridge, K.M., ..& Hayes, B.J. 2012. High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genetics*,13:53.

Ruiz, L., Bacigalupe, R., García-Carral, C., Boix-Amoros, A. Argüello, H., Silva, C.B.,… & Rodríguez J.M. 2019. Microbiota of human precolostrum and its potential role as a source of bacteria to the infant mouth. *Scientific Reports*, 6:37565. doi:10.1038/srep37565 1

Salih, D. A., El Hussein, A. M. & Singla L. D. 2014. Diagnostic approaches for tick-borne haemoparasitic diseases in livestock. *Journal of Veterinary Medicine and Animal Health*, 7(2), pp. 45-56. DOI: 10.5897/JVMAH2014. 0345

Salipante, S J. Kawashima, T., Rosenthal, C., Hoogestraat, D.R., Cummings, L.A., Sengupta, D.J., … & Hoffman, N.G. 2014. Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA based bacterial community profiling. *Applied and Environmental Microbiology Journal*, 80, 7583–7591.

Sanger, F., Nicklen, S., Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74: 5463–7.

Schisterman, E.F. & Vexler, A. 2008. To pool or not to pool, from whether to when: applications of pooling to biospecimens subject to a limit of detection. *Paediatric and Perinatal Epidemiology*, *22*(5), pp.486-496.

Shah, N., Tang, H., Doak, T.G. & Ye Y. 2011. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Biocomputing,* 165-176.

Shabbir MZ, Uvas LP, Ahmad A, Zahid MN, Nazir J, Nawaz M, et al. Sample collection guide. 2013 [cited 20 March 2019] Available from: http://www.nexusacademicpublishers.com/uploads/books/20140116135637.pdf

Singh, J., Behal A., Singla, N., Joshi, A., Birbian, N., Singh, … & Batra, N. 2009. Metagenomics: Concept, methodology, ecological inference and recent advances. *Biotechnology Journal*, 4, 480–494.

Sparagano, O.A.E., Allsopp, M.T.E.P., Mank, R.A., Rijpkema, S.G.T., Figueroa, J.V. & Jongejan, F. 1999. Molecular detection of pathogen DNA in ticks (Acari: Ixodidae): A review. *Experimental Applied Acarology,* 23: 929-960.

Taponen, S., McGuinness, D., Hiitiö, H., Simojoki, H., Zadoks, R. & Pyörälä, S. 2019. Bovine milk microbiome: a more complex issue than expected. *Veterinary Research*, 50:44.

Thomas, T., Gilbert, J., Meyer, F. 2012. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2 (3).

Tschopp, R., Abera, B., Sourou, S.Y., Guerne-Bleich, E., Aseffa, A., Wubete, A., …& Young, D. 2013. Bovine tuberculosis and brucellosis prevalence in cattle from selected milk cooperatives in Arsi zone, Oromia region, Ethiopia. *BMC Veterinary Research*, 9:163.

University of Minnesota Veterinary Diagnostic Laboratory. *Collecting Milk Samples* http://www.vdl.umn.edu/services-and-fees/udder-health-mastitis/collecting-milksamples/index.htm Date of access 20 February. 2019.

University Veterinarian & Animal Resources, *SOP: Blood Collection in Cattle* https://ouv.vt.edu/content/dam/ouv_vt_edu/sops/large-animal/sop-bovine-blood-collection.pdf Date of access 27 February. 2019.

Velmurugan G., Dinakaran, V. Rajendhran, J. & Swaminathan, K. 2020. Blood Microbiota and Circulating Microbial Metabolites in Diabetes and Cardiovascular Disease. *Trends in Endocrinology & Metabolism*, 20 (20):3.

Vidal, S., Kegler, K., Posthaus, H., Perreten, V. & Rodriguez-Campos, S. 2017. Amplicon sequencing of bacterial microbiota in abortion material from cattle. *Veterinary Research*, 48(1), pp.1-15.

Vientós-Plotts, A. I., Ericsson, A. C., Rindt, H., Grobman, M. E., Graham, A., Bishop, K., Cohn & L.A. Reinero, C.R. 2017. Dynamic changes of the respiratory microbiota and its relationship to fecal and blood microbiota in healthy young cats. *PLoS ONE*. 12:e0173818. doi: 10.1371/journal.pone.0173818

Vijayvargiya, P., Jeraldo, P.R., Thoendel, M.J., Greenwood-Quaintance, K.E., Garrigos, Z.E., Sohail, M.R., … & Patel, R. 2019. Application of metagenomics shotgun sequencing to detect vector-borne pathogens in clinical blood samples. *PLoS ONE*, 14(10): e0222915. https://doi.org/10.1371/journal. pone.0222915

Wang, Y., Song, J., Zhai, Y., Zhang, C., Gerritsen, J., Wang, H., ... & Ruan, Z. 2015. *Romboutsia sedimentorum* sp. nov., isolated from an alkaline-saline lake sediment and emended description of the genus *Romboutsia*. *International Journal of Systematic and Evolutionary Microbiology*, *65*(Pt_4), pp.1193-1198.

Wang, Y., Zhang, H., Zhu, L., Xu, Y., Liu, N., Sun, X., … & Zhu, R. 2018. Dynamic distribution of gut microbiota in goats at different ages and health states. *Frontiers in Microbiology*, 9, p.2509.

Wani, S.A., Bhat, M.A., Qureshi, S. & Bachh, A.S. 2003. Isolation of *Rhodococcus equi* from clinical bovine mastitis – A case report, *Indian Journal of Animal Research*, 37 (1), pp. 75-76.

Watanabe, N., Kryukov, K., Nakagawa, S., Takeuchi, J.S., Takeshita, M., Kirimura, Y., … & Inoue S. 2018. Detection of pathogenic bacteria in the blood from sepsis patients using 16S rRNA gene amplicon sequencing analysis. *PLoS ONE*, 13(8): e0202049. https://doi.org/10.1371/journal.pone.0202049

Weiner, M., Iwaniak, W. and Szulowski, K., 2011. Comparison of PCR-based AMOS, Bruce-Ladder and MLVA assays for typing of *Brucella* species. *Bull Vet Inst Pulawy*, *55*, pp.625-630.

Whittle, E., Leonard, M.O., Harrison, R., Gant, T.W., & Tonge, D.P. 2018. Multi-method characterization of the human circulating microbiome. *Frontiers in Microbiology*, 9:3266. doi: 10.3389/fmicb.2018.03266

Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. *Springer-Verlag,* New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

Witkowski, L., Rzewuska, M., Takai, S., Kizerwetter-Świda, M. & Kita, J. 2016. Molecular epidemiology of *Rhodococcus equi* in slaughtered swine, cattle and horses in Poland. *BMC Microbiology*, 16(1), pp.1-6.

Wooley, J.C. & Ye, Y. 2010. Metagenomics: Facts and artifacts, and computational challenges. *Journal of Computer Science and Technology*, 25(1): 71–81. doi:10.1007/s11390-010-9306-4

Wu, Y., Zhou, X., Zhang, X., Niu, H., Lyu, L., Liang, C., …& Jiang, S. 2021. Breast milk flora plays an important role in infantile eczema: cohort study in Northeast China. *Journal of Applied Microbiology*.

Young, C.C., Kämpfer, P., Chen, W.M., Yen, W.S., Arun, A.B., Lai, W.A., … & Chou, J.H. 2007. *Luteimonas composti* sp. nov., a moderately thermophilic bacterium isolated from food waste. *International Journal of Systematic and Evolutionary Microbiology*, *57*(4), pp.741-744.

Young, W., Hine B.C., Wallace O.A.M., Callaghan M. & Bibiloni R. 2015. Transfer of intestinal bacterial components to mammary secretions in the cow. *PeerJ.,* 3:e888; DOI10.7717/peerj.888

Zeineldin, M., Barakat, R., Elolimy, A., Salem, A., Elghandour M.M.Y., Monroy, C.J. 2018. Synergetic action between the rumen microbiota and bovine health, *Microbial Pathogenesis*. doi: 10.1016/j.micpath.2018.08.038

Zhang, R., Huo, W., Zhu, W. & Mao, S. 2015. Characterization of bacterial community of raw milk from dairy cows during subacute ruminal acidosis challenge by high-throughput sequencing. *Journal of the Science of Food and Agriculture*, 95(5), pp.1072-1079.

Zhou, W., Xu, H., Zhan, L., Lu, X. & Zhang, L. 2019. Dynamic development of fecal microbiome during the progression of diabetes mellitus in Zucker diabetic fatty rats. *Frontiers in Microbiology*, 10, p.232.

ANNEXURES

Annexure A

Phylum	Class	Order	Family	Genus	Species
Acidobacteriota	ABY1	0319-6G20	67-14	*1174-901-12*	*composti*
Actinobacteriota	Acidimicrobiia	Absconditabacteriales_(SR1)	A4b	*Acetitomaculum*	*necrophorum*
Aquificota	Acidobacteriae	Acetobacterales	Acetobacteraceae	*Acetivibrio*	*sedimentorum*
Armatimonadota	Actinobacteria	Acholeplasmatales	Acholeplasmataceae	*Acetoanaerobium*	
Bacteroidota	Alphaproteobacteria	Acidaminococcales	Acidaminobacteraceae	*Acholeplasma*	
Bdellovibrionota	Anaerolineae	Acidobacteriales	Acidaminococcaceae	*Acidaminobacter*	
Campilobacterota	Babeliae	Actinomarinales	Acidobacteriaceae_(Subgroup_1)	*Acidibacter*	
Chloroflexi	Bacilli	Actinomycetales	Actinomycetaceae	*Acidiphilium*	
Chrysiogenetota	Bacteroidia	Aeromonadales	Aerococcaceae	*Acidovorax*	
Cloacimonadota	Bdellovibrionia	Alteromonadales	Aeromonadaceae	*Acinetobacter*	
Cyanobacteria	Berkelbacteria	Anaerolineales	AKIW781	*Actinotalea*	
Deinococcota	Blastocatellia	Ardenticatenales	Akkermansiaceae	*Adhaeribacter*	
Dependentiae	BRH-c20a	Azospirillales	AKYG1722	*Adlercreutzia*	
Desulfobacterota	Campylobacteria	Babeliales	Alcaligenaceae	*ADurb.Bin063-2*	
Elusimicrobiota	Chloroflexia	Bacillales	Alcanivoracaceae	*Aeriscardovia*	
Fibrobacterota	Chrysiogenetes	Bacteriovoracales	Alteromonadaceae	*Aerococcus*	
Firmicutes	Cloacimonadia	Bacteroidales	Amb-16S-1323	*Aeromicrobium*	
Fusobacteriota	Clostridia	Bdellovibrionales	Anaerofustaceae	*Aeromonas*	
Gemmatimonadota	Coriobacteriia	Bifidobacteriales	Anaerolineaceae	*Agathobacter*	
Myxococcota	Cyanobacteriia	Blastocatellales	Anaeromyxobacteraceae	*Aggregicoccus*	
NB1-j	Dehalococcoidia	Bradymonadales	Anaerovoracaceae	*Akkermansia*	
Nitrospirota	Deinococci	Bryobacterales	Anaplasmataceae	*Alishewanella*	
Patescibacteria	Desulfitobacteriia	Burkholderiales	Arcobacteraceae	*Alistipes*	

Planctomycetota	Desulfobacteria	Caldicoprobacterales	Ardenticatenaceae	*Alkanindiges*
Proteobacteria	Desulfobulbia	Caldilineales	Atopobiaceae	*Allochromatium*
SAR324_clade(Marine_group_B)	Desulfovibrionia	Campylobacterales	Azospirillaceae	*Alloprevotella*
Spirochaetota	Desulfuromonadia	Candidatus_Chisholmbacteria	Bacillaceae	*Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium*
Sumerlaeota	Dojkabacteria	Candidatus_Falkowbacteria	Bacteriovoracaceae	*Altererythrobacter*
Synergistota	Elusimicrobia	Candidatus_Kuenenbacteria	Bacteroidaceae	*Alterococcus*
Verrucomicrobiota	Endomicrobia	Candidatus_Moranbacteria	Bacteroidales_RF16_group	*Alysiella*
	Fibrobacteria	Candidatus_Nomurabacteria	Bacteroidales_UCG-001	*Amaricoccus*
	Fusobacteriia	Candidatus_Peregrinibacteria	Bacteroidetes_vadinHA17	*Aminobacter*
	Gammaproteobacteria	Candidatus_Woesebacteria	Barnesiellaceae	*Anaerocella*
	Gemmatimonadetes	Caulobacterales	Bdellovibrionaceae	*Anaerofustis*
	Gitt-GS-136	CCD24	Beijerinckiaceae	*Anaerolinea*
	Gracilibacteria	Cellvibrionales	Bifidobacteriaceae	*Anaeromyxobacter*
	Holophagae	Chitinophagales	BIrii41	*Anaeroplasma*
	Ignavibacteria	Chloroflexales	Blastocatellaceae	*Anaerosolibacter*
	Incertae_Sedis	Chloroplast	Brevibacteriaceae	*Anaerosporobacter*
	Kiritimatiellae	Christensenellales	Bryobacteraceae	*Anaerostipes*
	Kryptonia	Chromatiales	Burkholderiaceae	*Anaerovorax*
	Lentisphaeria	Chrysiogenales	Butyricicoccaceae	*Anaplasma*
	Leptospirae	Chthoniobacterales	Caldicoprobacteraceae	*Anoxybacillus*
	Limnochordia	Cloacimonadales	Caldilineaceae	*Aquabacterium*
	Longimicrobia	Clostridia_or	Campylobacteraceae	*Aquimonas*
	Microgenomatia	Clostridia_UCG-014	Carnobacteriaceae	*Arcobacter*
	MVP-15	Clostridia_vadinBB60_group	Caulobacteraceae	*Arthrobacter*
	Myxococcia	Clostridiales	Cellulomonadaceae	*Atopobium*
	Negativicutes	Coriobacteriales	Cellvibrionaceae	*Atopococcus*
	Nitrospiria	Corynebacteriales	Chitinibacteraceae	*Azoarcus*
	OLB14	Cyanobacteriales	Chitinophagaceae	*Azovibrio*
	Oligoflexia	Cytophagales	Christensenellaceae	*Bacillus*

Omnitrophia	Defluviicoccales	Chromatiaceae	*Bacteroides*
Parcubacteria	Deinococcales	Chroococcidiopsaceae	*Bartonella*
Phycisphaerae	Desulfatiglandales	Chrysiogenaceae	*BBMC-4*
Planctomycetes	Desulfitobacteriales	Chthoniobacteraceae	*BD1-7_clade*
Polyangia	Desulfobulbales	Cloacimonadaceae	*Bergeyella*
Rhodothermia	Desulfovibrionales	Clostridiaceae	*Bibersteinia*
S0134_terrestrial_group	Desulfuromonadia_or	Coleofasciculaceae	*Bifidobacterium*
Saccharimonadia	DTU014	Comamonadaceae	*BIyi10*
SJA-28	Elusimicrobiales	Coriobacteriaceae	*Blastocatella*
Spirochaetia	Endomicrobiales	Corynebacteriaceae	*Blastomonas*
Sumerlaeia	Enterobacterales	Crocinitomicaceae	*Blastopirellula*
Synergistia	EPR3968-O8a-Bc78	Cryomorphaceae	*Blautia*
Syntrophia	Erysipelotrichales	Cyclobacteriaceae	*Bradyrhizobium*
Syntrophomonadia	Eubacteriales	D05-2	*Brevibacterium*
Syntrophorhabdia	Exiguobacterales	Defluviicoccaceae	*Brevundimonas*
Thermoanaerobaculia	Fibrobacterales	Defluviitaleaceae	*Brucella*
Thermoleophilia	Flavobacteriales	Deinococcaceae	*Bryobacter*
TK10	Fusobacteriales	Desulfatiglandaceae	*Burkholderia-Caballeronia-Paraburkholderia*
Vampirivibrionia	Gaiellales	Desulfobulbaceae	*Butyrivibrio*
Verrucomicrobiae	Gammaproteobacteria_Incertae_Sedis	Desulfocapsaceae	*Caldicoprobacter*
Vicinamibacteria	Gastranaerophilales	Desulfomicrobiaceae	*Caldilinea*
WWE3	Gemmatales	Desulfovibrionaceae	*Campylobacter*
	Gemmatimonadales	DEV007	*Candidatus_Cloacimonas*
	Haliangiales	Devosiaceae	*Candidatus_Finniella*
	Holophagales	Dietziaceae	*Candidatus_Omnitrophus*
	Ignavibacteriales	Dysgonomonadaceae	*Candidatus_Saccharimonas*
	Isosphaerales	Eggerthellaceae	*Candidatus_Soleaferrea*
	Izemoplasmatales	Elusimicrobiaceae	*Caproiciproducens*
	Kallotenuales	Endomicrobiaceae	*Caviibacter*

Kineosporiales	Enterobacteriaceae	*Cellulomonas*
Kiritimatiellales	Enterococcaceae	*Cellulosilyticum*
Kryptoniales	env.OPS_17	*Cerasicoccus*
Lachnospirales	Erysipelatoclostridiaceae	*Cetobacterium*
Lactobacillales	Erysipelotrichaceae	*Christensenellaceae_R-7_group*
Legionellales	Ethanoligenenaceae	*Chryseobacterium*
Leptospirales	Eubacteriaceae	*Chryseolinea*
Limnochordia_or	Exiguobacteraceae	*Chrysiogenes*
Lineage_IV	F082	*Citricoccus*
Longimicrobiales	Fibrobacteraceae	*Clostridium_sensu_stricto_1*
Methylococcales	Fibrobacterales_fa	*Colidextribacter*
Micrococcales	Flavobacteriaceae	*Comamonas*
Micromonosporales	Fusibacteraceae	*Conchiformibius*
Microtrichales	Fusobacteriaceae	*Conexibacter*
Monoglobales	Gemellaceae	*Coprococcus*
Mycoplasmatales	Gemmataceae	*Corynebacterium*
Myxococcales	Gemmatimonadaceae	*CPla-4_termite_group*
Nannocystales	Geothermobacteraceae	*Crinalium_SAG_22.89*
Nitrospirales	Gimesiaceae	*Curtobacterium*
Oceanospirillales	GZKB124	*Curvibacter*
Oligoflexales	Haliangiaceae	*Cutibacterium*
Oligosphaerales	Halomonadaceae	*Defluviicoccus*
Omnitrophales	Herpetosiphonaceae	*Defluviitaleaceae_UCG-011*
OPB41	Holophagaceae	*Deinococcus*
Opitutales	Hungateiclostridiaceae	*Delftia*
Oscillospirales	Hymenobacteraceae	*Denitrobacterium*
Paenibacillales	Hyphomicrobiaceae	*Desulfatiglans*
Paracaedibacterales	Iamiaceae	*Desulfobulbus*
Pasteurellales	Ilumatobacteraceae	*Desulfofustis*
Pedosphaerales	Intrasporangiaceae	*Desulfomicrobium*
Peptococcales	Isosphaeraceae	*Desulfonispora*

Peptostreptococcales-Tissierellales	Izemoplasmataceae	*DEV114*
Phycisphaerales	JG30-KF-CM45	*dgA-11_gut_group*
Pirellulales	Kineosporiaceae	*Dielma*
Planctomycetales	Kiritimatiellaceae	*Dietzia*
Polyangiales	Lachnospiraceae	*DNF00809*
Propionibacteriales	Lactobacillaceae	*Dokdonella*
Pseudomonadales	Legionellaceae	*Dorea*
Pseudonocardiales	Lentimicrobiaceae	*Draconibacterium*
Pyrinomonadales	Leptospiraceae	*Ehrlichia*
R7C24	Leptotrichiaceae	*Ellin6055*
RBG-13-54-9	Leuconostocaceae	*Elusimicrobium*
RF39	Limnochordia_fa	*EMP-G18*
Rhizobiales	Longimicrobiaceae	*Endomicrobium*
Rhodobacterales	LWQ8	*Enhydrobacter*
Rhodospirillales	M2PB4-65_termite_group	*Enorma*
Rhodothermales	Marinifilaceae	*Ensifer*
Rickettsiales	Marinilabiliaceae	*Enterobacter*
S085	Methylococcaceae	*Enterococcus*
Saccharimonadales	Methyloligellaceae	*Enterorhabdus*
SBR1031	Methylomonadaceae	*Ercella*
SJA-15	MgMjR-022	*Erysipelatoclostridium*
Solirubrobacterales	Microbacteriaceae	*Erysipelothrix*
Sphingobacteriales	Micrococcaceae	*Erysipelotrichaceae_UCG-009*
Sphingomonadales	Micromonosporaceae	*Escherichia/Shigella*
Spirochaetales	Microscillaceae	*EUB33-2*
Staphylococcales	Microtrichaceae	*Eubacterium*
Steroidobacterales	Mitochondria	*Exiguobacterium*
Streptomycetales	Monoglobaceae	*Facklamia*
Subgroup_7	Moraxellaceae	*Family_XIII_AD3011_group*
Sumerlaeales	MSB-3C8	*Family_XIII_UCG-001*
Synergistales	Muribaculaceae	*FD2005*

Syntrophales
Syntrophomonadales
Syntrophorhabdales
Tepidisphaerales
Thermoanaerobaculales
Thermomicrobiales
vadinBA26
Veillonellales-Selenomonadales
Verrucomicrobiales
Vicinamibacterales
Victivallales
WCHB1-41
Xanthomonadales

Mycobacteriaceae
Mycoplasmataceae
Myxococcaceae
Nannocystaceae
Neisseriaceae
Nitrosomonadaceae
Nitrospiraceae
Nocardiaceae
Nocardioidaceae
Oligosphaeraceae
Omnitrophaceae
Opitutaceae
Oscillatoriaceae
Oscillospiraceae
Oscillospirales_fa
Oxalobacteraceae
p-251-o5
p-2534-18B5_gut_group
Paenibacillaceae
Paludibacteraceae
Paracaedibacteraceae
Pasteurellaceae
Pedosphaeraceae
Peptococcaceae
Peptostreptococcaceae
Peptostreptococcales-Tissierellales_fa
PHOS-HE36
Phycisphaeraceae
Pirellulaceae
Planococcaceae
Porphyromonadaceae

Fermentimonas
Fibrobacter
Flaviaesturariibacter
Flavisolibacter
Flavobacterium
Flavonifractor
Flectobacillus
Flexilinea
Floricoccus
Fluviicola
Formivibrio
Frisingicoccus
Fusibacter
Fusobacterium
Gardnerella
GCA-900066575
Gemella
Gemmata
Gemmatimonas
Gemmobacter
Geobacillus
Geothermobacter
Giesbergeria
Granulicella
Guggenheimella
GWE2-31-10
Haliangium
Halomonas
Helcococcus
Herbaspirillum
Herbinix

Prevotellaceae	*Herpetosiphon*
Prolixibacteraceae	*HN-HF0106*
Propionibacteriaceae	*Holdemania*
Pseudohongiellaceae	*Howardella*
Pseudomonadaceae	*Hydrogenispora*
Pseudonocardiaceae	*Hydrogenoanaerobacterium*
Puniceicoccaceae	*Hydrogenophaga*
Pyrinomonadaceae	*Hymenobacter*
Rhizobiaceae	*Iamia*
Rhodanobacteraceae	*Ilumatobacter*
Rhodobacteraceae	*Incertae_Sedis*
Rhodocyclaceae	*Intestinimonas*
Rhodomicrobiaceae	*Janibacter*
Rhodothermaceae	*Jeotgalicoccus*
Rikenellaceae	*JGI-0000079-D21*
Rubinisphaeraceae	*Ketobacter*
Rubritaleaceae	*Klebsiella*
Ruminococcaceae	*Kocuria*
Saccharimonadaceae	*Krasilnikovia*
Saccharospirillaceae	*Lachnoclostridium*
Sandaracinaceae	*Lachnospiraceae_FCS020_group*
Saprospiraceae	*Lachnospiraceae_NK3A20_group*
SB-5	*Lachnospiraceae_NK4A136_group*
SC-I-84	*Lachnospiraceae_NK4B4_group*
Sedimentibacteraceae	*Lachnospiraceae_UCG-004*
Smithellaceae	*Lachnospiraceae_UCG-008*
Solirubrobacteraceae	*Lachnospiraceae_UCG-010*
Sphingobacteriaceae	*Lachnospiraceae_XPB1014_group*
Sphingomonadaceae	*Lactobacillus*
Spirochaetaceae	*Lautropia*
Spirosomaceae	*Lawsonella*
Spongiibacteraceae	*LD29*

Sporomusaceae	*Legionella*
Staphylococcaceae	*Lentimicrobium*
Steroidobacteraceae	*Leptolinea*
Streptococcaceae	*Leucobacter*
Streptomycetaceae	*Leuconostoc*
Succinivibrionaceae	*Limnobacter*
Sulfurimonadaceae	*Luteimonas*
Sulfurospirillaceae	*Luteitalea*
Sulfurovaceae	*Luteolibacter*
Sumerlaeaceae	*Lysinibacillus*
Sutterellaceae	*Lysobacter*
Synergistaceae	*Macellibacteroides*
Syntrophaceae	*Mageibacillus*
Syntrophomonadaceae	*Mailhella*
Syntrophorhabdaceae	*Mannheimia*
Tannerellaceae	*Marmoricola*
Tepidisphaeraceae	*Marvinbryantia*
Terrimicrobiaceae	*Massilia*
Thermoanaerobaculaceae	*Methylobacterium-Methylorubrum*
Thermotaleaceae	*Methylocaldum*
TRA3-20	*Methylocella*
Trueperaceae	*Methylomonas*
UCG-010	*Microbacterium*
UCG-011	*Micrococcus*
Unknown_Family	*Microcoleus_SAG_1449-1a*
vadinBE97	*mle1-7*
Veillonellaceae	*Mogibacterium*
Vermiphilaceae	*Monoglobus*
Verrucomicrobiaceae	*Moraxella*
Vicinamibacteraceae	*Mucilaginibacter*
Victivallaceae	*Mycobacterium*
WD2101_soil_group	*Mycoplasma*

Weeksellaceae	*Negativibacillus*
Woeseiaceae	*Niabella*
Xanthobacteraceae	*Nitrobacter*
Xanthomonadaceae	*Nitrosomonas*
Yersiniaceae	*Nitrospira*
	Niveispirillum
	NK4A214_group
	Nocardioides
	Noviherbaspirillum
	Oceanobacter
	Odoribacter
	Ohtaekwangia
	OLB13
	Olsenella
	OM27_clade
	Opitutus
	Ornithinicoccus
	Ornithinimicrobium
	Oscillibacter
	Oscillospira
	p-1088-a5_gut_group
	Paenibacillus
	Paeniclostridium
	Paludibacter
	Paludicola
	Papillibacter
	Parabacteroides
	Paracoccus
	Paraperlucidibaca
	Parapusillimonas
	Parasutterella
	Parvimonas

Patulibacter
Pedobacter
Pedomicrobium
Peptoclostridium
Peptostreptococcus
Peredibacter
Persicitalea
Phaeodactylibacter
Phascolarctobacterium
Phenylobacterium
Phyllobacterium
Pir4_lineage
Pirellula
Planctomicrobium
Planktothricoides_SR001
Planomicrobium
Pluralibacter
pLW-20
Polymorphobacter
Pontibacter
Porphyromonas
Prevotella
Prevotellaceae_Ga6A1_group
Prevotellaceae_UCG-001
Prevotellaceae_UCG-003
Prevotellaceae_UCG-004
Proteiniclasticum
Proteiniphilum
Pseudarcobacter
Pseudobacteroides
Pseudoflavonifractor
Pseudofulvimonas

Pseudomonas
Pseudonocardia
Pseudorhodoplanes
Psychrobacillus
Psychrobacter
Psychroglaciecola
Pygmaiobacter
Qipengyuania
Quadrisphaera
R76-B128
Raineyella
Ralstonia
Ramlibacter
RB41
RBG-16-49-21
Rhodobacter
Rhodococcus
Rhodomicrobium
Rhodopirellula
Rhodopseudomonas
Rhodovulum
Rikenella
Rikenellaceae_RC9_gut_group
Romboutsia
Roseburia
Roseimarinus
Roseimaritima
Roseisolibacter
Roseomonas
Rothia
Rubrivirga
Ruminiclostridium

Ruminobacter
Ruminococcus
Ruminofilibacter
S5-A14a
Saccharofermentans
Sandaracinus
Sanguibacteroides
Schlegelella
Sedimentibacter
Sediminispirochaeta
Serratia
SH-PL14
Silanimonas
Slackia
SM1A02
Smithella
Soehngenia
Solibacillus
Solirubrobacter
Solobacterium
Sphaerochaeta
Sphingobium
Sphingomonas
Sphingopyxis
Spirochaeta_2
Sporacetigenium
Sporobacter
Sporosarcina
Staphylococcus
Stenotrophobacter
Stenotrophomonas
Steroidobacter

Streptococcus
Subgroup_10
Sulfuricurvum
Sulfurospirillum
Sulfurovum
Sumerlaea
Syner-01
Syntrophococcus
Syntrophomonas
Syntrophorhabdus
Syntrophus
Tepidibacter
Tepidisphaera
Termite_planctomycete_cluster
Terriglobus
Terrimicrobium
Terrisporobacter
Tessaracoccus
Tetrasphaera
Thauera
Tissierella
TM7a
TM7x
Treponema
Tropicimonas
Truepera
Trueperella
Tundrisphaera
Turicibacter
Tuzzerella
UCG-002
UCG-004

UCG-005
UCG-007
UCG-009
UCG-012
Veillonella
Vicingus
Wenxinia
Woeseia
XBB1006
Zavarzinella

Annexure B

Faeces only	Milk only	Blood only	Blood_Faeces_Milk Intersection	Blood_Milk Intersection	Faeces_Milk Intersection	Blood_Faeces Intersection
Adlercreutzia	*1174-901-12*	*Anaplasma*	*Acetitomaculum*	*Aerococcus*	*Atopobium*	*Arthrobacter*
Aeriscardovia	*Acetivibrio*	*Bartonella*	*Acinetobacter*	*Anaeroplasma*	*Blautia*	*Prevotella*
Denitrobacterium	*Acetoanaerobium*	*Ehrlichia*	*Agathobacter*	*Anaerosporobacter*	*Clostridium_sensu_stricto_1*	*Flexilinea*
Eubacterium	*Acholeplasma*	*Sporosarcina*	*Akkermansia*	*Aquabacterium*	*Corynebacterium*	*Colidextribacter*

Lachnospiraceae_NK4A136_group	*Acidaminobacter*	*Paenibacillus*	*Alistipes*	*Bradyrhizobium*	*CPla-4_termite_group*	*Termite_plancto mycete_cluster*
Lachnospiraceae_NK4B4_group	*Acidibacter*	*Enterobacter*	*Alloprevotella*	*Burkholderia-Caballeronia-Paraburkholderia*	*DNF00809*	*Paludicola*
Papillibacter	*Acidiphilium*	*Pluralibacter*	*Anaerovorax*	*Butyrivibrio*	*Erysipelotrichaceae_UCG-009*	
Planomicrobium	*Acidovorax*	*Curtobacterium*	*Bacillus*	*Campylobacter*	*GCA-900066575*	
Pseudoflavonifractor	*Actinotalea*	*Enterococcus*	*Bacteroides*	*Caproiciproducens*	*Lachnospiraceae_UCG-010*	
Psychrobacter	*Adhaeribacter*	*Holdemania*	*Candidatus_Saccharimonas*	*Cerasicoccus*	*Lysinibacillus*	
Syntrophococcus	*ADurb.Bin063-2*	*Amaricoccus*	*Candidatus_Soleaferrea*	*Cutibacterium*	*Mageibacillus*	
Tepidibacter	*Aeromicrobium*	*Caldicoprobacter*	*Christensenellaceae_R-7_group*	*Dielma*	*Mogibacterium*	
UCG-012	*Aeromonas*	*Desulfonispora*	*Coprococcus*	*EMP-G18*	*Parabacteroides*	
XBB1006	*Aggregicoccus*	*Slackia*	*Defluviitaleaceae_UCG-011*	*Enhydrobacter*	*Psychrobacillus*	
	Alishewanella		*dgA-11_gut_group*	*Escherichia/Shigella*	*Pygmaiobacter*	
	Alkanindiges		*Dorea*	*Gardnerella*	*Rikenella*	
	Allochromatium		*Elusimicrobium*	*GWE2-31-10*	*Saccharofermentans*	
	Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium		*Enterorhabdus*	*Hydrogenoanaerobacterium*	*Solobacterium*	
	Altererythrobacter		*Erysipelatoclostridium*	*Klebsiella*	*Sporacetigenium*	
	Alterococcus		*Family_XIII_AD3011_group*	*Lachnospiraceae_UCG-004*	*Terrisporobacter*	
	Alysiella		*Family_XIII_UCG-001*	*Lentimicrobium*		
	Aminobacter		*Fibrobacter*	*Marvinbryantia*		
	Anaerocella		*Flavonifractor*	*Mycoplasma*		
	Anaerofustis		*Fusobacterium*	*Negativibacillus*		
	Anaerolinea		*Incertae_Sedis*	*Nitrobacter*		
	Anaeromyxobacter		*Lachnoclostridium*	*Oscillospira*		
	Anaerosolibacter		*Lachnospiraceae_FCS020_group*	*Paracoccus*		

Anaerostipes	*Lachnospiraceae_NK3A20_group*	*Parasutterella*
Anoxybacillus	*Lachnospiraceae_UCG-008*	*Peptostreptococcus*
Aquimonas	*Mailhella*	*Phyllobacterium*
Arcobacter	*Monoglobus*	*Porphyromonas*
Atopococcus	*NK4A214_group*	*Ralstonia*
Azoarcus	*Odoribacter*	*Rhodococcus*
Azovibrio	*Olsenella*	*Ruminobacter*
BBMC-4	*Oscillibacter*	*Sandaracinus*
BD1-7_clade	*p-1088-a5_gut_group*	*Sediminispirochaeta*
Bergeyella	*Paeniclostridium*	*Serratia*
Bibersteinia	*Phascolarctobacterium*	*Sphaerochaeta*
Bifidobacterium	*Pirellula*	*Sphingomonas*
Blyi10	*Prevotellaceae_Ga6A1_group*	*Staphylococcus*
Blastocatella	*Prevotellaceae_UCG-001*	*Stenotrophomonas*
Blastomonas	*Prevotellaceae_UCG-003*	*Tuzzerella*
Blastopirellula	*Prevotellaceae_UCG-004*	
Brevibacterium	*Pseudomonas*	
Brevundimonas	*Rikenellaceae_RC9_gut_group*	
Brucella	*Romboutsia*	
Bryobacter	*Roseburia*	
Caldilinea	*Ruminiclostridium*	
Candidatus_Cloacimonas	*Ruminococcus*	
Candidatus_Finniella	*Sanguibacteroides*	
Candidatus_Omnitrophus	*Solibacillus*	
Caviibacter	*Streptococcus*	
Cellulomonas	*Treponema*	

	Cellulosilyticum	*Turicibacter*
	Cetobacterium	*UCG-002*
	Chryseobacterium	*UCG-004*
	Chryseolinea	*UCG-005*
	Chrysiogenes	*UCG-009*
	Citricoccus	
	Comamonas	
	Conchiformibius	
	Conexibacter	
	Crinalium_SAG_22.89	
	Curvibacter	
	Defluviicoccus	
	Deinococcus	
	Delftia	
	Desulfatiglans	
	Desulfobulbus	
	Desulfofustis	
	Desulfomicrobium	
	DEV114	
	Dietzia	
	Dokdonella	
	Draconibacterium	
	Ellin6055	
	Endomicrobium	
	Enorma	
	Ensifer	
	Ercella	
	Erysipelothrix	
	EUB33-2	

Exiguobacterium
Facklamia
FD2005
Fermentimonas
Flaviaesturariibacter
Flavisolibacter
Flavobacterium
Flectobacillus
Floricoccus
Fluviicola
Formivibrio
Frisingicoccus
Fusibacter
Gemella
Gemmata
Gemmatimonas
Gemmobacter
Geobacillus
Geothermobacter
Giesbergeria
Granulicella
Guggenheimella
Haliangium
Halomonas
Helcococcus
Herbaspirillum
Herbinix
Herpetosiphon
HN-HF0106
Howardella

Hydrogenispora
Hydrogenophaga
Hymenobacter
Iamia
Ilumatobacter
Intestinimonas
Janibacter
Jeotgalicoccus
JGI-0000079-D21
Ketobacter
Kocuria
Krasilnikovia
Lachnospiraceae_XPB1014_group
Lactobacillus
Lautropia
Lawsonella
LD29
Legionella
Leptolinea
Leucobacter
Leuconostoc
Limnobacter
Luteimonas
Luteitalea
Luteolibacter
Lysobacter
Macellibacteroides
Mannheimia
Marmoricola
Massilia

Methylobacterium-Methylorubrum
Methylocaldum
Methylocella
Methylomonas
Microbacterium
Micrococcus
Microcoleus_SAG_1449-1a
mle1-7
Moraxella
Mucilaginibacter
Mycobacterium
Niabella
Nitrosomonas
Nitrospira
Niveispirillum
Nocardioides
Noviherbaspirillum
Oceanobacter
Ohtaekwangia
OLB13
OM27_clade
Opitutus
Ornithinicoccus
Ornithinimicrobium
Paludibacter
Paraperlucidibaca
Parapusillimonas
Parvimonas
Patulibacter
Pedobacter

Pedomicrobium
Peptoclostridium
Peredibacter
Persicitalea
Phaeodactylibacter
Phenylobacterium
Pir4_lineage
Planctomicrobium
Planktothricoides_SR001
pLW-20
Polymorphobacter
Pontibacter
Proteiniclasticum
Proteiniphilum
Pseudarcobacter
Pseudobacteroides
Pseudofulvimonas
Pseudonocardia
Pseudorhodoplanes
Psychroglaciecola
Qipengyuania
Quadrisphaera
R76-B128
Raineyella
Ramlibacter
RB41
RBG-16-49-21
Rhodobacter
Rhodomicrobium
Rhodopirellula

Rhodopseudomonas
Rhodovulum
Roseimarinus
Roseimaritima
Roseisolibacter
Roseomonas
Rothia
Rubrivirga
Ruminofilibacter
S5-A14a
Schlegelella
Sedimentibacter
SH-PL14
Silanimonas
SM1A02
Smithella
Soehngenia
Solirubrobacter
Sphingobium
Sphingopyxis
Spirochaeta_2
Sporobacter
Stenotrophobacter
Steroidobacter
Subgroup_10
Sulfuricurvum
Sulfurospirillum
Sulfurovum
Sumerlaea

	Syner-01
	Syntrophomonas
	Syntrophorhabdus
	Syntrophus
	Tepidisphaera
	Terriglobus
	Terrimicrobium
	Tessaracoccus
	Tetrasphaera
	Thauera
	Tissierella
	TM7a
	TM7x
	Tropicimonas
	Truepera
	Trueperella
	Tundrisphaera
	UCG-007
	Veillonella
	Vicingus
	Wenxinia
	Woeseia
	Zavarzinella