Elucidating the role of the rumen microbiome in cattle feed efficiency and its potential as a reservoir for novel enzyme discovery

by

André Luis Alves Neves

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

Doctor of Philosophy

in

Animal Science

Department of Agricultural, Food and Nutritional Science
University of Alberta

Abstract

The rapid advances in omics technologies have led to a tremendous progress in our understanding of the rumen microbiome and its influence on cattle feed efficiency. However, significant gaps remain in the literature concerning the driving forces that influence the relationship between the rumen microbiota and host individual variation, and how their interactive effects on animal productivity contribute to the identification of cattle with improved feed efficiency. Furthermore, little is known about the impact of mRNA-based metatranscriptomics on the analysis of rumen taxonomic profiles, and a strategy for the discovery of lignocellulolytic enzymes through the targeted functional profiling of carbohydrate-active enzymes (CAZymes) remains to be developed. Study 1 investigated the dynamics of rumen microorganisms in cattle raised under different feeding regimens (forage vs. grain) and studied the relationship among the abundance of these microorganisms, host individuality and the diet. To examine host individual variation in the rumen microbial abundance following dietary switches, hosts were grouped based on the magnitude of microbial population shift using $\log_2$-fold change ($\log_2$-fc) in the copy numbers of bacteria, archaea, protozoa and fungi. Three groups of $\log_2$-fc in the bacterial and fungal abundance (Low, $\log_2\text{-fc} < -1$; Stable, $-1 < \log_2\text{-fc} < 1$; and High, $\log_2\text{-fc} > 1$) were identified from the magnitude of change in baseline rumen microbial populations. By monitoring the microbial population shift within the same animal in response to the diet, significant ecological features of rumen microorganisms were identified and shed new light on their dynamic roles in animal feed utilization and individual variation. Study 2 compared the outcomes of two methods, Kraken (mRNA based) and a pipeline developed in-house based on Mothur (16S rRNA based), concerning the taxonomic

profiles (bacteria and archaea) of rumen microbial communities using total RNA sequencing of rumen fluid collected from cattle with different feed conversion ratios (FCR). Both approaches revealed a similar phyla distribution of the most abundant taxa, with Bacteroidetes, Firmicutes, and Proteobacteria accounting for approximately 80% of total bacterial abundance. For bacterial taxa, although 69 genera were commonly detected by both methods, an additional 159 genera were exclusively identified by Kraken. Kraken detected 423 species, while Mothur was not able to assign bacterial sequences to the species level. For archaea, both methods generated similar results only for the abundance of Methanomassiliicoccaceae and *Methanobrevibacter ruminantium*. Although Kraken enhanced the microbial classification at the species level, identification of bacteria or archaea in the rumen was limited due to a lack of reference genomes for the rumen microbiome. Study 3 investigated the effect of cattle breeds on specific ruminal taxonomic microbial groups and functions associated with FCR, using two genetically related Angus breeds as a model. Total RNA was extracted from rumen content samples collected from purebred Black and Red Angus bulls fed the same forage diet and then subjected to metatranscriptomic analysis. Multivariate discriminant analysis (sPLS-DA) and analysis of composition of microbiomes (ANCOM) were conducted to identify microbial signatures characterizing Black and Red Angus cattle. Although Black and Red Angus are genetically similar, sPLS-DA detected 25 bacterial species and ten functions that differentiated the rumen microbial signatures between those two breeds. ANCOM revealed an association between FCR and breed with *Chitinophaga pinensis* and *Clostridium stercorarium*, suggesting that these bacterial species may play a key role in the feed conversion efficiency of forage-fed bulls. Study 4 combined selective pressure

to enrich the rumen for lignocellulolytic microbes with bioinformatic tools to guide the discovery of unknown CAZymes in the microbiome. It was demonstrated that the rumen microbiome increased the abundance of lignocellulolytic bacteria, such as *Fibrobacter succinogens*, and a diverse set of CAZymes over time, including 18 uncharacterized members of the family GH11 (xylanases) and three of the family GH45 (endoglucanases). Further experiments confirmed the lignocellulolytic activity of xylanase using such approach. In summary, the data presented in this thesis provide fundamental knowledge on the role of the rumen microbiome in cattle feed efficiency and offers opportunities to further explore the potential of the rumen as a source for novel enzyme discovery.

Preface

This thesis is an original work by Andre Luis Alves Neves and is part of a collaborative project between Profs. Leluo Guan at the University of Alberta and Prof. Kim H. Ominski of the University of Manitoba. The Veterinary Services and the Animal Care Committee of the University of Manitoba granted ethical approval for all experimental procedures described in Chapters 2, 3, 4 and 5.

Chapter 2 of this thesis has been submitted to FEMS Microbiology Ecology as "Dynamics of microbial populations driven by interactions between diet and host shed light on individualized rumen microbiota" by Andre L. A. Neves, Yanghong Chen, Eoin O'Hara, Tim McAllister, Kim H. Ominski and Le Luo Guan. (2019). ALAN was responsible for laboratory and data analysis, as well as manuscript writing. YC performed the laboratory analysis, while LLG, KHO, TM, and EOH revised and contributed to the writing of the manuscript.

Chapter 3 of the thesis has been published as Andre L. A. Neves, Fuyong Li, Bibaswan Ghoshal, Tim McAllister and Le Luo (2017). "Enhancing the resolution of rumen microbial classification from metatranscriptomic data using Kraken and Mothur". Frontiers Microbiology; 8: 2445. doi: 10.3389/fmicb.2017.02445. ALAN and BG executed Kraken, and FL executed the pipeline based on Mothur. ALAN analyzed the data and performed the statistical analysis. ALAN, FL and BG executed the experiment and wrote the manuscript. LLG and TM contributed to the experiment and revised the manuscript.

Chapter 4 has been submitted to Animal as "Taxonomic and functional assessment reveals the effect of Angus breed genetics on rumen microbial signatures" by Andre L. A. Neves, Yanhong Chen, Kim-Anh L. Cao, Siddhartha Mandal, Thomas J. Sharpton, Tim McAllister, and Le Luo Guan. ALAN run Kraken, analyzed the data, performed the statistical analysis, and wrote the manuscript. KALC and SM helped ALAN run the sPLS-DA and ANCOM, respectively. TJS helped ALAN run ShotMAP. YC assisted with the laboratory analysis. LLG and TM contributed to the manuscript writing.

Dedication

This thesis is dedicated to my Lord and Savior Jesus Christ, who said "*I am the resurrection, and the life: he that believeth in me, though he were dead, yet shall he live. And whosoever liveth and believeth in me shall never die. Believe thou this?*", **KJV Bible, John 11.25,26.**

I sincerely appreciate your support and guidance: I love you my **Lord Jesus**!

This thesis is also dedicated to my beautiful wife, Euzilene Neves, and our beloved son, Daniel Neves!

Thanks Lene for the patience and encouragement during this time, I love you and Daniel!

Acknowledgements

First, I would like to express my gratitude to my supervisor Dr. Le Luo Guan for her dedication, mentoring, advice and patience during my Ph.D. program. I am proud to say to everyone that I had an excellent supervisor, who provided me with the best opportunities to learn and grow in various aspects of my professional career and personal life. You may have no idea about the important role you played in my life, Dr. Guan! Throughout these four years, you helped me get rid of all academic fears, and I am passionate about science because you taught me so. Thanks for accepting me as your student in May/2015 and for making me what I am now. I will take your academic and life teachings for the rest of my days! Thank you very much!

I also want to thank the funding bodies AITF for the doctorate scholarship and EMBRAPA for the financial support during the program.

To my committee members: Drs. Tim McAllister, Michael Steele, and Paul Stothard. Thanks for providing me with constructive suggestions on my projects that contributed to my professional development. Additionally, I want to thank Dr. McAllister for introducing me to Dr. Guan in 2015 and for providing me with resources to improve my English skills when I started writing my first paper in English in 2015. Since then, I have written in a much faster pace! I also appreciated being a student of Drs. Steele and Stothard. I enjoyed all the classes at the UofA, especially those at DRTC!

I thank Dr. Lynn McMullen and Dr. Carl Yeoman for taking their time to review my thesis and agreeing to be my arms in length and external examiners.

I also thank the R developers and those involved in the elaboration of bioinformatic software for their dedication and provision of amazing tools that were vastly employed in this thesis. Special thanks to Drs. Kim-Anh L. Cao, Siddhartha Mandal, and Thomas J. Sharpton for their collaborations in the statistical and bioinformatic approaches implemented in the Chapters 3 and 4 of this thesis. Also a special thanks to my dear friend Dr. Eóin O'Hara for his critical review of the findings described in this thesis, especially Chapters 2 and 5.

I also thank Prof. Joanne Lemieux for opening the doors of her lab to a humble student like me and for helping me perfect the protein work! To Dr. Elena Arutyunova, for

her incredible help and guidance in the protein purification and enzyme assays. Thanks also to my dear brother Jiangkun Yu for being such an incredible friend and great support 24/7! Also thanks to my sister Marisol for joining Jiangkun and me in the journey to study the proteins. I also thank Yanhong for her incredible help in all lab analysis of this thesis and for being such a great friend and a beautiful human being.

I also would like to thank my little sister Yang Song, my preferred officemate of all times and one of my best friends, as well as Biba, Anusha, Cameron and Fuyong for the friendship and help with my Ph.D. program.

I want to thank all graduate students and staff who are unforgettable people: Urmila, Yan Meng, Ou, Ke, Nilu, Huizeng, Bin, Junhong, Charnpal, Rebecca, Guanxiang, Mimi, Xiao Xie, Jitka, Zhi, Junhua, Tao, Yawei, Duanqin, Diming, Xu Sun, Officemates: Andrea and Andre, and the College Plaza people: Robert, Tiago, Huaigang (“the dark cut man”), Jiyuan, Anahid and Janelle for the friendship and friendly conversations.

I would like to express my gratitude to my brothers and sisters at the West Edmonton Christian Assembly, especially those of the Ignite group, for encouraging me during my doctorate: Pastors Jane, Sandra, and J. Hayford and their families, brothers Hayford, Olu, Noel, Mark, Greg, Tim, Art, Marco, sisters Wendy, Linda, Marlene, Joyce, Cindy. I also thank the members of my home church in Brazil for the prayers and encouragement, especially Pastor Luiz Oliveira, sister Benigna, brother Alex, and the Assembly of God at the Aracaju’s airport neighborhood.

I would like to thank Humberto Reis, Sava Maria, Mariu, Cristiane, Mariazinha, Jose Araujo, Duda, Cau, Coda and their entire families, Marilene, Eldio, Edilene, Reginaldo, Keniel, Ezinho, Manu, Nilzete, Maria Pereira, Marcela, Eloi, Ivan, Mari, and Tereza for the friendship, encouragement and support from Brazil and Portugal!

If I could say thanks to someone else, I would say thanks to my grandparents Zeca and Noemia and my dear aunt Ana Maria for supporting my education since the very beginning and for being great examples of perseverance.

Last but not least, I would like to express my gratitude to my father Jorge Neves and my mother Doraci Neves, my brothers Anderson and Anselmo and their families, as well as my sister Erica Neves, for the encouragement of all times and for believing in my dreams. I love you forever!

Table of Contents

List of Tables

List of Figures

Abbreviations

3-NOP – 3-Nitrooxypropanol

ANCOM – Analysis of composition of microbiomes

ANOVA – Analysis of variance

CAZymes - Carbohydrate-active enzymes

CH_4 – Methane

CO_2 – Carbon dioxide

DMI – Dry matter intake

DGGE – Denaturing gradient gel electrophoresis

FAO – Food and Agriculture Organization of the United Nations

FCR – Feed conversion ratio

FDR – False discovery rate

FE – Feed efficiency

GH – Glycoside hydrolase

GHG – Greenhouse gas

GWP – Global warming potential

H_2 – Hydrogen

IPCC – United Nations Intergovernmental Panel on Climate Change

ITS – Internal transcribed spacer gene

MixMC - Multivariate Statistical Framework to Gain Insight into Microbial Communities

NDF – Neutral detergent fiber

OTU – Operational taxonomic unit

PCA – Principle component analysis

PCR – Polymerase chain reaction

PICRUSt – Phylogenetic investigation of communities by reconstruction of unobserved states

QIIME – Quantitative insights into molecular ecology

qPCR – Quantitative polymerase chain reaction

RFI – Residual feed intake

RIM-DB – Rumen and intestinal methanogens database

RIN – RNA integrity number

rRNA/DNA – ribosomal RNA/DNA gene

sPLS-DA – Sparse partial least square regression – discriminant analysis

TMM –Trimmed mean of m-values

VFA – Volatile fatty acids

Chapter 1

Literature review

1.1 Introduction

The agri-food industry contributes to the economic development of many countries through the provision of employment, income and food security. It has been estimated that this industry employs over 1 billion people around the world, accounting for 1 in 3 of all workers of the global workforce (Network 2016). In the domestic scenario, the beef industry is the largest component of the Canadian food processing sector, with annual sales surpassing $28 billion, including exports exceeding $6 billion and direct employment for over 66,000 people (Statistics Canada, 2017). The Canadian beef sector is the 12th largest globally and provides around 1.9% (around 62 million metric tonnes) of the world's beef supply (USDA, 2017). However, increased global demand for food will continue to rise as the global population will continue to increase to 9.15 billion people by the year 2050. Despite the importance of Canada's beef industry in the world, the dramatic increase in the human population will require a further growth of 70% in food production to meet the global demands for adequate nutrition (Nations 2015, Alexandratos and Bruinsma 2012).

However, with the increase in the demand for animal protein, there is a simultaneous increase in greenhouse gases (GHG) produced by the environmental footprint of ruminant agriculture (Huws et al. 2018). On a global scale, it has been shown that ruminants contribute between 9 and 11% of total anthropogenic GHG production (Pickering et al. 2015b), with approximately 44% of ruminant emissions generated in the form of methane (Rojas-Downing et al. 2017). Furthermore, methane emissions represent 2-12% energy loss from the gross energy intake in cattle, and this outcome reduces host feed efficiency (Johnson and Johnson 1995). Therefore, mitigating methane emissions from ruminants is a necessary step in the process of reducing the negative impacts of GHG on the environment, and also to improve animal feed efficiency.

One fascinating microbial community that has raised much interest is the one that resides in the upper digestive tract of ruminants, termed the rumen microbiome (Shaani et al. 2018). A detailed understanding of the structure and function of the rumen microbiome is central to improving animal feed efficiency and minimizing energy loss as a result of methane emissions, as the rumen microbiota plays a fundamental role in the food digestion and provision of energy for the host (Matthews et al. 2018). In the last decades, the advent of high-throughput sequencing technologies has greatly advanced our understanding of the major microbial populations and functional pathways of the rumen microbiome involved in methane emissions and feed efficiency (Denman et al. 2018, Huws et al. 2018). However, little is known about the precise mechanisms of host-microbe interactions in ruminants, and how the dynamics of individual changes in the rumen microbial communities impact the host feed efficiency. Since inter-individual variability and the microbiome are likely to vary simultaneously (Bashiardes et al. 2018), a comprehensive understanding of their interactive effects on the host may create opportunities to enhance animal productivity and alleviate the pressure of the livestock sector on the environment.

In addition to its role in feed degradation and host energy provision, the rumen microbiome is a valuable source of novel enzymes with major applications in the biotechnology and biofuels industries (Seshadri et al. 2018, Ribeiro et al. 2016). The rumen microbiome remains a source of valuable bioactives for the biotechnology industry due to the proven ability of the rumen enzymes to break down and release the vast energy stored in the most abundant carbon polymer on the planet, namely lignocellulose (Hess et al. 2011, Gharechahi and Salekdeh 2018, Meng Qi et al. 2011). The importance of discovering novel rumen-derived enzymes capable of converting lignocellulose into biobased alternatives stems from the growing demands for a bioeconomy less dependent on petroleum-derived fuels (Clark et al. 2006). Despite the wealth of knowledge on the rumen microbiome and its relevance as a source for enzyme discovery, limitations related to the analysis of microbiome datasets still exist and they are exacerbated by the vast differences in analytical approaches employed across studies. Thus, standardization of the sample processing workflow - from collection to analysis - to allow more reliable comparisons of results across studies will provide comprehensive insights into the rumen

microbial composition and its functional capacity, which will ultimately lead to the development of intervention strategies to reduce livestock methane emissions and to discover new enzymes.

This literature review consists of four main sections. The first will summarize the rumen microbiome and its microbes, and the contribution of the rumen microbiome to ruminant feed efficiency and methane emissions. The second section relates to the rumen microbiome as a reservoir for novel enzyme discovery. The third section reviews molecular approaches used to investigate the rumen microbiome and to discover rumen enzymes. The fourth section discusses the optimal statistical approaches for analyzing microbiome datasets generated from the molecular technologies.

1.2 Importance of the rumen microbiome to cattle production

1.2.1 Rumen microbial diversity and functions

The ruminant forestomach is composed of four compartments, the rumen, reticulum, omasum, and abomasum. The rumen is the largest compartment and contains a vast array of anaerobic microorganisms (bacteria, archaea, protists, and fungi) that functions in a coordinated fashion to ferment feed into volatile fatty acids (VFA), which are ultimately used as energy source for milk and meat production (Russell and Rychlik 2001). The rumen is the primary site of microbial fermentation and is lined with a filiform and foliate epithelium with an extended surface area that facilitates the absorption of VFAs (Steele et al. 2016). Temperature, pH, buffering capacity, osmotic pressure, and redox potential (Eh) contribute to the maintenance of ruminal homeostasis and create the ideal habitat for the growth of anaerobic and facultative anaerobic microbes (Russell and Rychlik 2001, Weimer 1992). A distinguishing feature of the rumen microbiome compared to other microbiomes (e.g., soil) is that it exhibits a highly anaerobic condition (Eh = -150 to -350 mV) that creates the ideal environment for microbial fermentation of feed particles (Russell and Rychlik 2001). This fermentative activity results in a continuous production of gases that are largely removed via eructation (Russell and Rychlik 2001, Weimer 1992). Chemical reactions with reducing agents and a dedicated population of facultative anaerobic microorganisms maintain the anaerobic conditions in the rumen for the whole microbial population by scavenging oxygen taken in with food and drinking water (Weimer

1992). This synergistic microbial consortium is made up of cellulolytic, hemicellulolytic, amylolytic, and biohydrogenating species that exhibit a high level of functional redundancy (Firkins and Yu 2015, Hobson and Stewart 1997, Weimer 1992). Enhancing our understanding of this complex microbial community is critical to provide the framework needed to address the current challenges (e.g., mitigating methane emissions and improving feed efficiency) of the ruminant agriculture chain. The next sections will discuss in depth the major microbial members of the rumen microbiome (bacteria, archaea, protozoa, and fungi) and their functions. A more in-depth discussion about the rumen microbiome membership and their functions can be found in Huws et al. (2018) and Matthews et al. (2018).

1.2.1.1 Rumen bacteria

Bacteria represent the most abundant organisms of the rumen microbiome (density of 10^{10}-10^{11} cells/ml rumen fluid), making up more than 50% of the cell mass (Creevey et al. 2014). Rumen bacteria are dominated by members of the Firmicutes, Bacteroidetes and Proteobacteria phyla, which can account for >90% of the total bacterial population (M. S. McCabe et al. 2015, Henderson et al. 2015, Fouts et al. 2012). Bacteria can be free in the rumen fluid, attached to ingested feed particles (firmly or loosely), adhered to the rumen wall (known as epimural bacteria), or associated with eukaryotic organisms (Stewart et al. 1988, Cheng et al. 1979, Miron et al. 2001, McAllister et al. 1994a). Numerically, the digesta-associated bacteria are the most abundant among the four fractions (McAllister et al. 1994a), followed by the liquid-associated (planktonic) fraction of the rumen fluid, which comprises around 30% of the rumen bacterial population (Millen et al, 2016). The digesta-associated bacteria ferment feed or utilize the end-products of this fermentation, and are dominated by fermentative species. In contrast, the epimural bacteria are often facultative anaerobes, producing urease and scavenging oxygen to assist in the maintenance of the ruminal anaerobic environment (Liu et al. 2016).

Early work indicated that at least 22 bacterial species had been characterized by the beginning of the 21st century using culture dependent-methods (Russell and Rychlik 2001), but molecular techniques showed that less than 20% of the rumen microbiota can

be cultured using standard culture-based techniques (Krause et al. 2013, Nocker et al. 2007). The assessment of the rumen bacterial diversity in the Ribosomal Database project revealed 13,478 bacterial sequences and the estimated number of bacterial species in the rumen was of approximately 7,000 species (Kim et al. 2011). Recently, a global census of rumen bacterial membership analyzed 742 rumen content samples collected from 32 ruminant species across 35 countries and found that 30 dominant bacterial genera were present in over 90% of animals and represented 89.4% of all generated sequence data (Henderson et al. 2015). This study concluded that the current technologies have already defined the dominant groups of rumen bacteria, and also revealed the existence of a "core" group of rumen microbes, ubiquitous to ruminants worldwide, containing the most abundant ruminal bacterial species, such as *Prevotella spp., Butyrivibrio spp.,* and *Ruminococcus spp.* (Henderson et al. 2015). However, the taxonomy of ruminal bacteria is continually being updated, and the possibility of discovering unculturable or previously unrecognized ruminal bacteria still exists.

1.2.1.2 Rumen archaea

The domain archaea is divided into two different kingdoms; Euryarchaeota, consisting of methanogens and extreme halophytes, and Crenarchaeota, consisting of hyperthermophiles and nonthermophiles (Bayley et al. 1999). Rumen methanogens belong exclusively to Euryarchaeota (with a range of 10^6 to 10^8 cells per ml) and account for 0.3-3.3% of the ruminal microbial population (Peter H. Janssen and Marek Kirs 2008, Lin et al. 1997). These microbes require a very low redox potential (Eh = -300 mV) to grow and are among the strictest anaerobes on Earth (Sirohi et al. 2010). The assessment of the rumen archaeal community present in the Ribosomal Database project revealed 3,516 archaeal sequences and an estimated number of 1,500 species (Kim et al. 2011). Despite this diversity, studies have shown that 90% of rumen archaea belong to the *Methanobrevibacter* (63.2% of the methanogenic population), *Methanomicrobium* (7.7%), *Methanosphaera* (9.8%), Rumen Cluster C (now referred as Methanomassiliicoccaceae, 7.4%), and *Methanobacterium* (1.2%) genera (Peter H. Janssen and Marek Kirs 2008, Patra et al. 2017). Most methanogens (e.g., *Methanobrevibacter* species) use hydrogen gas as electron donors to reduce CO_2 to methane (formate can also be used as an

electron donor and may contribute to the production of up to 18% of ruminal methane) (Hungate 1967). Other species, such as *Methanosphaera stadtmanae*, only produce methane through the reduction of methanol with hydrogen, having one of the strictest energy metabolisms of all methanogenic archaea (Peter H. Janssen and Marek Kirs 2008). The continuous production of methane carried out by methanogens leads to a low concentration of hydrogen in the rumen and creates a favorable environment for the growth of other species and more efficient fermentation (Peter H. Janssen and Marek Kirs 2008).

However, methane produced in the rumen is eructed and released into the environment, leading to atmospheric pollution (Matthews et al. 2018). Due to these environmental concerns, research on rumen methanogens has attracted great interest in the last decade (Wallace et al. 2014, Johnson and Johnson 1995). Therefore, manipulating the rumen microbiome to mitigate the activity of methanogens would help reduce the negative impact of methane emissions on the environment and improve the feed efficiency of ruminants. Strategies of mitigating methane emissions in ruminants and the contribution of the rumen microbiome to host feed efficiency are discussed further in sections 1.2.2 and 1.2.3.

1.2.1.3 Rumen protozoa

Protozoa are found in the range of 10^4 to 10^6 cells per ml in the rumen and are accountable for 20-50% of the rumen microbial biomass (McSweeney and Mackie 2012). They are involved in lipid hydrolysis and can produce large amounts of hydrogen via their hydrogenosomes (Tymensen et al. 2012), and thus may contribute to methane production via interspecies hydrogen transfer with methanogens (Hobson and Stewart 1997, S. Kittelmann et al. 2015, Gijzen et al. 1988). Morphological studies have identified over 250 ciliate species in a range of ruminant hosts, represented by around 40 genera (Veira 1986, Williams and Coleman 1997). The *Entodinium* genus is the most abundant protozoan in the rumen, and other common genera include *Polyplastron, Epidinium* and *Eudiplodinium* (Kittelmann et al. 2013a, Carberry et al. 2012a, Sylvester et al. 2004). Protozoa rapidly digest feed particles via engulfment, and also predate on bacteria and other small microbes, to ultimately convert them to an iodophilic storage polymer through a vast array

of enzymes such as amylases and glucosidases (Matthews et al. 2018, McSweeney and Mackie 2012).

The contribution of protozoa to rumen fermentation remains controversial in the scientific community. Deliberate removal of protozoa from the rumen, known as defaunation, does not have a detrimental effect on the animal survival but it can affect feed degradation efficiency in the rumen (Newbold et al. 2015). Defaunation can be carried out by several techniques such as chemical removal of protozoa through the use of copper sulphate, calcium peroxide, alcohol ethoxylate, coconut oil, linseed oil or soya oil hydrolysate (Jouany 1995). A recent meta-analysis analyzed 23 *in vivo* defaunation studies with the goal of determining the function of rumen protozoa and concluded that the removal of protozoa from the rumen decreased the degradation of neutral and acid detergent fibers, indicating that protozoa play important roles in fiber digestion (Newbold et al. 2015). In addition to their functional importance as fibrolytic microbes, protozoa have been associated with methanogenesis since defaunation reduces methane output by 11% (Morgavi et al. 2012b, Newbold et al. 2015). Protozoa-associated methanogens also account for approximately 37% of methanogenesis in ruminants (Finlay et al. 1994), and defaunation could disrupt this symbiotic relationship in order to reduce methane production (Matthews et al. 2018). However, research outcomes suggest that a strategy to eliminate all protozoa in the rumen may not be the best approach to mitigate methane emissions as it can compromise the activity of those protozoa involved in fiber degradation (Newbold et al. 2015), and this can affect the ruminal fermentation stability and consequently host productivity.

1.2.1.4 Rumen fungi

Anaerobic fungi (10^3 to 10^6 zoospores/ml) were firstly described by Colin Orpin in 1975 who reported them in the gastrointestinal tract of herbivores, particularly in the rumen and caecum (Orpin 1975). These strictly anaerobic microbes within the phylum Neocallimastigomycota are key players in the degradation of lignocellulosic feedstuffs within the gut of ruminant and non-ruminant herbivores (Gruninger et al. 2014). The anaerobic rumen fungi are currently grouped into eight genera (*Neocallimastix*, *Piromyces, Ontomyces*, *Buwchfawromyces*, *Caecomyces*, *Orpinomyces*, *Anaeromyces*,

and *Cyllamyces*) characterized primarily using classical microscopy (Dollhofer et al. 2015). However, next-generation sequencing technology has revealed an immense and undiscovered biodiversity which may have been underappreciated in the past (Koetschan et al. 2014). Each genus is distinguishable by morphological features such as thallus morphology (monocentric vs. polycentric), rhizoids morphology (filamentous vs. bulbous) and zoospore flagellation (monoflagellate vs. polyflagellate) (Ho and Barr 1995, Ozkose et al. 2001).

To illustrate the importance of anaerobic fungi in plant cell wall degradation, several experiments provided insight into their contribution to fiber digestion, feed intake, rumen fermentation and rumen metabolism (Gruninger et al. 2014). For example, the removal of anaerobic fungi from the rumen can reduce voluntary feed intake and dry matter degradation, indicating that feed digestion is usually impaired when they are removed from the rumen (Akin et al. 1988, Morrison et al. 1990, Gordon and Phillips 1998). Furthermore, the elimination of anaerobic fungi from the rumen significantly reduced the degradation of dry matter, neutral detergent fiber, acid detergent fiber and the activity of carboxymethylcellulases (Ford et al. 1987, Gordon and Phillips 1993, Gao et al. 2013), further confirming the importance of these microbes for fiber digestion in the rumen.

1.2.2 Feed efficiency and the rumen microbiome

Feed efficiency is a measure used to determine the relative ability of cattle to convert feed into useable products (beef or milk), and it is a moderately heritable trait in cattle (Cammack et al. 2018, Berry and Crowley 2013). In the past, beef cattle performance was traditionally evaluated using traits that measured total output (i.e., weight gain or lean meat carcass yield) without considering input measurements such as feed intake. As the costs associated with feed increased dramatically to 55-70% of the total costs in beef cattle operations (NRC 2016), the selection for feed efficient cattle is essential for the continuous growth in farm profitability as well as environmental sustainability. Feed conversion ratio (FCR) (gain in body weight per unit of feed consumed) has been considered as one of the standard measures of feed efficiency for individual farmers, and selecting for low FCR (efficient) animals usually results in a corresponding improvement in gross efficiency (e.g., growth rate) (Korver 1988). The popularity of FCR among farmers

is a reflection of its moderate heritability for gross efficiency (Crews 2005), and this means that selecting for low FCR animals leads to a simultaneous improvement in gross efficiency.

However, FCR-based selection strategies result in increased growth rate and mature cow size, leading to increased nutritional requirements to maintain the breeding female that can thus increase the feeding costs (Gunsett 1984). To circumvent this limitation, a metrics called residual feed intake (RFI) has been proposed as an alternative to FCR (Koch et al. 1963). Expected feed intake is calculated by regressing average daily gain and metabolic mid-test weight on standardized feed intake, and the difference of this value from actual feed intake provides a “residual feed intake” or RFI value. Genetically independent of growth, animals may be classified as low-RFI (efficient) or high-RFI (inefficient) (Alemu et al. 2017, Kong et al. 2016, Carberry et al. 2014b, Montanholi et al. 2010). The main limitation to adopting the RFI methodology for predicting feed efficiency on a large-scale basis is the inability to accurately measure feed intake in a feedlot setting. However, with advancements in individual feed intake recording systems, such as GrowSafe® Feeding Systems (GrowSafe Systems Ltd., Airdrie, Alberta), accurate feed intake monitoring has been possible, and the use of RFI measures to estimate feed efficiency in cattle has increased in recent years (Thompson 2015).

Regardless of the differences in the feed efficiency metrics (FCR vs. RFI) used in beef cattle operations, it has been reported a significant variation in feed efficiency from one animal to the other that is influenced mainly by the feeding behavior, energy metabolism, the rumen microbiota composition, and the genetic background of the host (Cantalapiedra-Hijar et al. 2018). Although all these factors are usually interconnected with each other, studies have revealed that energy metabolism-related factors (at the cellular level) are more relevant to explain individual variations in feed efficiency than digestion-related factors (e.g., feeding behavior) (Cantalapiedra-Hijar et al. 2018). The main finding described in those studies is that feed efficient cattle produce less heat than inefficient animals due to a decreased protein turnover and a higher efficiency of mitochondrial ATP production generated from the digestion and catabolism of carbohydrates, proteins and lipids in the rumen (Kong et al. 2016, Cantalapiedra-Hijar et

al. 2018). The lower energy metabolic rate observed in feed efficient cattle is independent of changes in feeding behavior (e.g., voluntary feed intake), suggesting that metabolic energy-consuming metrics need to be taken into consideration to explain individual variations in ruminants' feed efficiency (Cantalapiedra-Hijar et al. 2018, Nkrumah et al. 2006).

Another important factor contributing to individual variations in feed efficiency is the rumen microbiota as it plays essential roles in feed degradation and host energy provision (Matthews et al. 2018). Several studies have shown linkages between rumen microbial profiles and feed efficiency (Shabat et al. 2016, Guan et al. 2008, Carberry et al. 2012b, Fuyong Li and Le Luo Guan 2017), including the linkages between efficient cattle and the abundances of the bacterial *Lachnospiraceae* and *Veillonellaceae* families (Myer et al. 2015a, F. Li and L. L. Guan 2017), and a number of archaeal taxa (Carberry et al. 2014a, Carberry et al. 2014b, F. Li and L. L. Guan 2017). In the attempt to understand the rumen microbiota of feed efficient cattle, a study of the functional activity of the rumen microbial population revealed that efficient cattle (Low RFI) had lower rumen microbial diversity and richness than inefficient cattle (High RFI) (Fuyong Li and Le Luo Guan 2017), demonstrating that the rumen microbiota of efficient animals is less complex and more specialized in harvesting energy from the diet than inefficient animals (Shabat et al. 2016). More recently, Li et al. (2019b) analyzed the rumen microbiota of 709 beef cattle and showed that multiple factors including sex, breed, and diet are responsible for variations in the rumen microbial composition. The authors found that the relative abundance of ~34% of microbial taxa as well as the copy number of total bacteria were associated with feed efficiency and had a heritability estimate ≥ 0.15, suggesting that they are influenced by the host genetic background.

As mentioned above, the interactions of the rumen microbiome with the host can lead to individual variations in feed efficiency and understanding their influence on animal production may provide opportunities to create strategies aiming to reduce the impact of ruminant agriculture on the environment as discussed in the next section.

1.2.3 Rumen microbial contribution to methane emissions

The rumen microbial fermentation represents a significant source of greenhouse gas emissions, with CH_4 production contributing to reduce host dietary energy availability by up to 12% (Johnson and Johnson 1995). This outcome has led to the urgent need for the development of strategies to mitigate methane production in the livestock sector while concurrently improving animal feed efficiency to meet the increasing global demand for food.

The development of strategies to mitigate methane production by ruminants depends on our understanding of the linkage between methanogens and methane production in the rumen. Hydrogen is a by-product of normal rumen fermentation, and is a regulator of methane production in ruminants (Hegarty et al. 2007), whereby methanogens consume ruminal H_2 in the terminal step of carbohydrate fermentation (Deppenmeier 2002). As described previously, hydrogenotrophic methanogenesis is the predominant pathway in the rumen, and is carried out mainly by *Methanobrevibacter* species, which typically account for over 90% of archaeal 16S rRNA reads (Hristov et al. 2012). Species of *Methanosphaera, Methanimicrococcus,* and *Methanobacterium* also utilize H_2 to produce CH_4 (Tapio et al. 2017). The less abundant methylotrophic methanogens include members of the *Methanosarcinales*, *Thermoplasmatales*, *Methanophaera,* and *Methanomassillicoccaceae* (Tapio et al. 2017, P. H. Janssen and M. Kirs 2008, Poulsen et al. 2013). Due to the important role played by rumen archaea in CH_4 production, several strategies have been developed to reduce enteric methane emissions, either by directly targeting the methanogen population or attempting to reduce the substrates for archaeal metabolism.

These mitigation strategies include dietary manipulation, plant lipid feeding, synthetic methanogen inhibitor supplementation, defaunation, and genetic selection for reduced methane emitting animals (Hristov et al. 2013, McAllister et al. 2015, Martin et al. 2010, Knapp et al. 2014, Kumar et al. 2014, Cammack et al. 2018, Pickering et al. 2015a). A promising compound for reducing ruminal methanogenesis is 3-nitrooxypropanol (3-NOP), which has been shown to reduce methane emissions by up to 30% (Hristov et al. 2015, Jayanegara et al. 2018). However, reductions in methane emissions do not always result in a redirection of energy towards enhancing animal

production efficiency while supplementing with 3-NOP, likely because the biochemical processes involved in VFA and H_2 production derived from cellulose breakdown requires energy input (Jayanegara et al. 2018). Thus, the combination of 3-NOP supplementation with phloroglucinol has been proposed to capture the excess of ruminal H_2 from methanogenesis and generate valuable metabolites (e.g., propionate) for the host (Martinez-Fernandez et al. 2017). Although promising, it should be noted that the costs of the application of 3-NOP supplementation and the long-term effect of this strategy on the abatement of methane production in ruminants still need to be assessed in more details.

Another promising method of reducing ruminal methane emissions per unit product is through selecting genetic traits that increase the general efficiency of production in ruminants (Kamke et al. 2016). This strategy aims at associating genetic traits (e.g., weaning weight, dag score, muscle depth, etc.) with the rumen microbiota in order to predict methane yield phenotypes across ruminant species (Kamke et al. 2016). Researchers have recently characterized the rumen microbiota of sheep and found that some microbial profiles were associated with low methane production in some individuals, which they named as low methane emitting sheep. Species closely related to the lactic acid producing bacteria *Sharpaea azabuensis* showed an increased abundance in low methane emitting sheep, likely indicating rapid heterofermentative pathways resulting from higher rumen turnover rates. However, H_2 producing *Ruminoccocaceae*, *Lachnospiraceae* and *Verricomicrobia* were more prevalent in the rumen of high methane yielding sheep (Kittelmann et al. 2014, Kamke et al. 2016). In other ruminant species like cattle and goats, high methane emitting animals exhibited an increased abundance of *Verrucomicrobia* and *Synergistetes* bacteria and a decreased abundance of methanogens (Denman et al. 2015, Wallace et al. 2015, Martinez-Fernandez et al. 2016). Collectively, these findings suggest that the selection for animals that yield less methane per unit of consumed feed based on the rumen microbiota may be beneficial for animal production systems.

It is important to note that in addition to concentrating efforts to mitigating methane emissions from ruminants and improving cattle feed efficiency through the understanding

of the rumen microbiota, researchers have discovered valuable enzymes in the rumen that can be applied in the biotechnology industry (e.g., feed additives), and this topic will be discussed in the next section.

1.3 Rumen microbiome: a potential reservoir of industrially important microbial enzymes

The success of ruminants in digesting recalcitrant biomass is largely explained by the ability of ruminal microorganisms to metabolize the constituents of plant structural carbohydrates (Russell and Rychlik 2001), as the ruminant genome lacks the genes encoding for the enzymes involved in the breakdown of lignocellulose (Seshadri et al. 2018). Structural carbohydrates (plant cell wall) are complex structures composed mainly of lignocellulose – the most abundant organic material on Earth – which is a matrix composed of polysaccharide networks, glycosylated proteins, and lignin (Chafe 1969, McNeil et al. 1984). Although abundant in nature, lignocellulose represents a critical barrier for the conversion of plant biomass into feed sources and biofuels owing to the lack of an efficient enzymatic system to deconstruct lignocellulose and release all fermentable sugars it contains (Bader et al. 2010).

The lignocellulolytic matrix is comprised of three main components: cellulose, hemicellulose, and lignin. Cellulose is a linear biopolymer of anhydroglucopyranose molecules, connected by β-1,4-glycosidic bonds and represents the most abundant component of the lignocellulose complex, accounting for approximately 20-30% of the dry weight of most plant primary cell walls (Chafe 1969, McNeil et al. 1984). Adjacent cellulose chains are bound by hydrogen bonds, hydrophobic interactions, and Van der Waal's forces that generate a parallel alignment of crystalline structures known as microfibrils (Zhang et al. 2006). The second most abundant component of lignocellulose is hemicellulose, which consists of mixed polymers of pentoses (including xylose and arabinose), hexoses (mainly mannose, less glucose, and galactose) and sugar acids (Bhatia et al. 2012). Although the composition of hemicelluloses varies considerably depending on the plant source (Saha 2003, Saha 2000), hemicellulose is composed mainly of xylans with a backbone structure of β-1,4-linked xylose residues attached to

various sidechain molecules such as acetic acid, coumaric acid, glucuronic acid, ferulic acid, and arabinose (McNeil et al. 1984). Lignin is the third main polymer of lignocellulose and possesses a structure comprised of three aromatic alcohols: coniferyl alcohol, sinapyl, and p-coumaryl (McNeil et al. 1984). Compared to cellulose and hemicellulose, lignin is the most recalcitrant structure for rumen microorganisms to degrade and extract metabolizable energy (Himmel et al. 2007, Sanchez 2009). The surface area available for microbial attachment and the retention time of the ingested feed in the rumen are listed as the main obstacles for lignin digestion by rumen microorganisms (Wang and McAllister 2002). Thus, solubilization of lignin is a necessary step for increasing digestibility of lignocellulosic compounds through microbial fermentation in the rumen (Weimer et al. 1990).

Due to the recalcitrant nature of lignocellulose, the degradation of the complex cross-linkage structure intertwined by cellulose, hemicellulose, and lignin requires the synchronized action of a diverse array of rumen microbial enzymes to cleave the numerous bounds within the plant cell wall structure and access specific substrates in the lignocellulosic fiber (Wang and McAllister 2002). Owing to the efficacy of the rumen enzymes to deconstruct lignocellulose, researchers have recommended these biocatalysts to the animal feed industry as feed additives for ruminants (Sehgal et al. 2008) and non-ruminants (poultry and swine) (Theodorou et al. 2007). Potent polysaccharide-degrading enzymes derived from the rumen have also raised the interest of several industries including brewing, food, textile, paper, and biofuel production (Gruninger et al. 2014). A broad group of enzymes present in the rumen which are of particular interest to the industry are the carbohydrate-active enzymes (CAZymes), and they will be discussed in the topic of the next section.

1.3.1 CAZymes

Microbes play a significant role in regulating the biochemical processes of feed digestion in the rumen (Russell and Rychlik 2001) and are valuable sources of enzymes that have various biotechnological applications (Seshadri et al. 2018). The interest in discovering new enzymes for the saccharification of lignocellulose has led to the creation of an extensive database describing CAZymes families of structurally-related catalytic and

carbohydrate-binding modules (CBMs) that are continually updated to cover all CAZymes across organisms and subfields of glycoscience (Cantarel et al. 2009, Lombard et al. 2014). CAZymes and their associated CBMs are classified based on sequence similarity and encompass biocatalysts whose job is to modify and cleave carbohydrates (glycoside hydrolyses (GHs), polysaccharide lyases - PLs, carbohydrate esterases - CEs) and synthesize them (glycosyltransferases) (Cantarel et al. 2009, Lombard et al. 2014). Today, well over 100 GHs families of over 50 CBM families have been described in the Carbohydrate-Active enZymes - CAZy database. In the next sections, important CAZymes will be discussed, with the reader being encouraged to visit the CAZy database for more details about the structure and classification of CAZymes (http://www.cazy.org).

1.3.1.1 Cellulases

Cellulases are family members of the broad group of GHs (e.g., GH5, GH6, GH9, GH45), which have gained interest for a number of biotechnological applications (e.g., treatment of paper pulp) owing to their ability to hydrolyze 1,4 β-D-glycosidic linkages of the cellulose chain (Bayer et al. 1998). Based on the structure and functionality, cellulases have been classified into three groups: 1) Endoglucanases, a group of cellulases that exhibits a deep cleft or groove to accommodate the cellulose chain at any point along its length in order to cleave internal bonds at amorphous sites of new chain ends; 2) Exoglucanases, in contrast, are a group of cellulases that possess the active site in an extended loop that forms a tunnel, though which one of the termini of a cellulose chain can be threaded; and 3) β-glucosidases, which are cellulases that hydrolyze cellobiose to generate two molecules of glucose, and are often associated with the microbial cell surface when cellodextrins are transported into the cell (Bhat and Bhat 1997).

In addition to the free enzymes, self-assembled multienzyme complexes known as cellulosomes have been reported in several anaerobic environments including the rumen (Artzi et al. 2016, Bayer et al. 1998). To date, the only rumen species that exhibits an elaborated cellulosomal system is the mesophilic bacterium *Ruminoccocus flavefaciens* (Dassa et al. 2014). The major difference between free enzymes and cellulosomal enzymes is that the free enzymes usually contain a CBM to guide the catalytic domain to

the substrate, while the cellulosomal enzymes carry a dockerin domain that incorporates the enzyme into the cellulosomal complex (Artzi et al. 2016). The attachment of the cellulosomal machinery to cellulose chains is achieved via cellulose-binding proteins (Family 3a-CBM) of the scaffoldin subunit, which contains one or more cohesin modules connected to other types of functional modules. The arrangement of the modules on the scaffoldin subunit and the specificity of the cohesins and/or dockerin for their modular counterpart dictate the overall structure of the cellulosome (Artzi et al. 2016, Bayer et al. 1998, Doi and Kosugi 2004). Due to their high hydrolytic activity, cellulosomes have a great potential for the degradation of biomass and are the focus of much effort to engineer an effective cellulosomal structure for the conversion of lignocellulose into valuable products, such as biofuels.

1.3.1.2 Hemicellulases

In contrast to cellulose degradation, the digestion of hemicelluloses poses a different challenge, as this group of polysaccharides includes widely different types of sugars or non-sugar constituents with different types of bonds. Thus, hemicellulases can be divided into two main groups: a) those that cleave the mainchain backbone (e.g., mannanases and xylanases), and b) those that degrade sidechain substituents or short end products (e.g., arabinofuranosidase) (Shallom and Shoham 2003). Thus, the catalytic modules of hemicellulases can be either glycosyl hydrolases that hydrolyze glycosidic bonds, or carbohydrate esterases, which hydrolyze ester linkages of acetate or ferulic acid side groups (Shallom and Shoham 2003). Due to the different structures of the hemicellulose molecule, details of the various hemicellulases needed for their catabolism are given below.

1.3.1.2.1 Mannanases

Mannan is a component of hemicellulose and consists of β-1,4 linkages between mannose monomers that form the hemicellulose cross-linkages (Hogg et al. 2003). β-Mannanases (e.g., GH5, GH26) hydrolyze mannan-based hemicelluloses and release short β-1,4-manno-oligomers, which can be further hydrolyzed to mannose by the action of β-mannosidases. Deficiency in these enzymes in ruminants (termed β-mannosidosis) causes skeletal abnormalities (Shallom and Shoham 2003), demonstrating the relevance

of β-mannases for the digestion of mannan in ruminants. Recently, researchers have applied exogenous β-mannanases from *Aspergillus niger* in the animal feed industry as an additive to improve feed conversion efficiency in dairy cattle (Tewoldebrhan et al. 2017). However, the rumen microbiome possesses the ability to degrade mannan as demonstrated in the discovery of a multifunctional glycosyl hydrolase encoded in the genome of the bacterium *Prevotella bryantti* B14 (Palackal et al. 2007).

1.3.1.2.2 Arabinofuranosidases

Arabinose is found in conjunction with xylan as a hemicellulose component of plant cell walls, with arabinose units being attached to xylan via alpha-1,2,1,3,1,5 or linked to C2 or C3 positions on the arabinoxylan chain (Shallom and Shoham 2003). In the rumen, arabinose units can be cleaved off the xylose backbone by arabinofuranosidases (e.g., GH3, GH43) expressed by rumen bacteria such as *Ruminococcus albus* (Greve et al. 1984).

1.3.1.2.3 Ferulic acid esterases

Ferulic acid esterase (e.g., CE1) is a group of enzymes that forms a subclass of carboxylic ester hydrolases. These enzymes hydrolyze the bonds between hydroxycinnamates and sugars to release ferulic acid (Rashamuse et al. 2007, Wang and McAllister 2002). In the rumen, these ester bonds are cleaved by ferulic acid esterases encoded in the genome of the rumen fungi *Anaeromyces mucronatus* (M. Qi et al. 2011) as well as by rumen bacteria such as *P. ruminicola* (Kabel et al. 2011).

1.3.1.2.4 p-Coumaric acid esterases

p-Coumaric acid esterase or p-coumaroyl esterase (e.g., CE1) is an essential enzyme for efficient degradation of lignocellulose biomass in the rumen (Wang and McAllister 2002). This enzyme targets the p-Coumaroyl ester bonds that connect lignin to hemicelluloses, releasing p-Coumaric acid. Surprisingly, this type of enzyme is exclusively produced by anaerobic rumen fungi (phylum Neocallimastigomycota) and was not yet described in rumen bacteria (Borneman et al. 1990), further strengthening the ecological role and significance of anaerobic fungi for deconstructing lignocellulose biomass in the rumen (Gruninger et al. 2014).

1.3.1.2.5 Xylanases

The xylan molecule consists of β-1,4 linked xylopyranosyl residues and contains sidechains with acetyl group and L-arabinofuranosyl residues. Xylanases (e.g., GH5, GH8, GH10, GH11, GH51) are responsible for the hydrolysis of xylan by breaking the glycosidic linkages present in the xylan backbone (Shallom and Shoham 2003). Like the cellulases, the xylanases can be classified into three groups: endoxylanases, β-xylosidases and acetyl xylan esterases (Beg et al. 2001). All three enzymes hydrolyze the xylan molecule, rendering D-xylose sugar (Kosugi et al. 2001). The scientific interest in discovering new xylanases is reflected by the vast number of research papers published in recent years describing numerous xylanases applications in the pulp and paper industries (Beg et al. 2001, El Enshasy et al. 2016), and as exogenous enzyme preparations marketed by the animal feed industry (Wang and McAllister 2002). The rumen harbors a wide range of microbes cable of degrading xylan, including *Prevotella* spp. (such as *P. ruminocola, P. albensis, P. brevis, and P. bryantii*) and non-*Prevotella* spp. (such as *Ruminococcus* and *Fibrobacter*) (Russell and Rychlik 2001, Dai et al. 2015).

1.3.1.3 Pectinases

Pectin exists in the primary cell wall and represents the plant's first line of defense against dehydration and penetration by phytopathogens. The pectin's structure is a backbone of alpha-1,4-linked residues of D-galacturonate that is degraded by rumen pectinolytic enzymes (e.g., PL11, GH28) including pectin lyases, polygalacturonases and pectin methylesterases (Wang and McAllister 2002). One of the major pectinolytic bacterial species that inhabits the rumen is *Lachnospira multiparus*, which produces pectin lyases and pectin methylesterases (Silley 1985, Russell and Rychlik 2001). In addition to that bacterial species, rumen fungi also exhibit pectinolytic enzymes (Orpin and Joblin 1997, Gordon and Phillips 1992).

1.3.1.4 Polyphenol degrading enzymes

Feed consumed by ruminants contain not only the nutrients required by the host animal for maintenance and production, but also hold naturally occurring plant secondary compounds such as tannins, saponins, phenolic acids and silica that usually cause

adverse effects on the activity of fibrolytic enzymes (McAllister et al. 1994b, Bae et al. 1997). However, some gastrointestinal microbes of ruminants are able to break down tannin-protein complexes through an enzyme known as tannin acyl hydrolase (tannase), which catalyzes the hydrolysis of ester bonds present in gallotannins, complex tannins, and gallic acid esters (Rodríguez et al. 2009, Ramírez et al. 2008). Tannase activity in ruminants has been reported mainly in *Streptococcus caprinus* (now *Streptococcus gallolyticus*) (Brooker et al. 1994) and *Selenomonas ruminantium* (Skene and Brooker 1995). These tannninolytic microbes possess tannin-degrading ability to tolerate tannins in feeds, and their selection to promote long-term protection against tannin toxicity can be used to improve the nutritive value of tannin-rich feeds (Goel et al. 2005).

It is important to note that a key component to improve our understanding of the rumen microbiome and its role in animal productivity and enzyme discovery is the development of bioinformatic tools specially tailored to overcome the technical challenges of the analysis of massively paralleled, high-throughput sequencing data. The section 1.4 of this chapter will discuss in depth the molecular biology techniques designed for investigating the rumen microbiome, its microbes and enzymes.

1.4 Molecular approaches to study the rumen microbiome

The work of Robert Hungate, outlined in the book “The Rumen and its Microbes” (Hungate 1966) formed the foundation for investigations of basic rumen microbial ecology in the context of agricultural production (McCann et al. 2017). Hungate’s roll-tube method called for the use of rumen fluid in growth media to isolate ruminal microbes, and quickly became the most common approach available for rumen microbiologists to discover new species (Bryant and Burkey 1953, Bryant and Robinson 1961). By the 1990s and before the rise of the ‘omics’ approaches, rumen microbiologists perfected culture media techniques to isolate rumen bacteria and redefined our understanding of how the anaerobic microbiota functions (Krause et al. 2013). This pioneering work enabled the identification of at least 22 major bacteria and improved the knowledge of the interrelationship between carbohydrate and protein-nitrogen metabolism driven by ruminal microbes (Russell and Rychlik 2001). From the use of the Hungate “roll-tube” technique, modern culture-based approaches were decisive to unveiling the biochemical and physiological activities of

rumen microbes since the cultured organisms present in the public collections can be studied both *in vitro* and *in vivo* (Huws et al. 2018).

Our ability to culture rumen microbes was further improved by technologies (e.g., dilution to extinction and removal of reducing agents) that can be used to culture the as yet unculturable rumen microbes (Kenters et al. 2011, Poelaert et al. 2017). Despite the success of cultivation experiments to characterize undescribed rumen microbes, most cultures are not available in every collection, and culture-based methods may also be limited to, or biased toward, strains that are highly abundant and organisms which are of specific interest to research (Zehavi et al. 2018, Huws et al. 2018). Furthermore, it has been demonstrated that the number of microbial species isolated and characterized from the rumen is low (Morgavi et al. 2013), highlighting the importance of molecular biology methods to overcome the limitations of culture-based approaches to study the rumen microbial composition and diversity (Morgavi et al. 2013). As new technology has become available, what initially involved the isolation and culture of strains in the laboratory, has now moved to large-scale sequencing of the 'total' detectable rumen microbiota nucleic acids (e.g., metagenomics and metatranscriptomics) (Denman et al. 2018). The comprehensive application of these new technologies in most ruminant microbiota studies has caused significant advances in our understanding of the rumen microbiota diversity and function, and therefore these techniques will be discussed in more details in the following sections.

1.4.1 Metataxonomics

Metataxonomics (or amplicon sequencing) refers to the high-throughput sequencing analysis of amplified taxonomic marker genes and has been used to characterize the taxonomic composition of microbiota of many ecosystems (Marchesi and Ravel 2015). In ruminant studies, this method has been routinely employed to capture variations in the microbial composition in response to perturbations such as dietary changes, subacute acidosis, and different feed efficiencies and methane production (Petri et al. 2013, McGovern et al. 2017, Henderson et al. 2015, Myer et al. 2015b). Studies of bacterial and archaeal diversity have relied on primer sets that target the 16S rDNA or rRNA gene (Deusch et al. 2015, Matthew Sean McCabe et al. 2015, McGovern et al. 2017), whereas

18S rRNA and Internal Transcribed Spacer (ITS) genes have been used for eukaryotic targets, typically protists and fungi (Kittelmann et al. 2013b, Sandra Kittelmann et al. 2015). While metataxonomics is a cheap, fast technology to characterize the rumen composition and diversity in a wide range of ruminant hosts (McGovern et al. 2017, Tapio et al. 2016, F. Li et al. 2016, M. S. McCabe et al. 2015, Myer et al. 2015a, Jami et al. 2013, Jami and Mizrahi 2012), it is subject to several limitations. These include PCR biases (primer specificity and sensitivity, non-specific annealing, differential amplification specificity of taxonomic groups, artefact formation), poor resolution beyond the genus level, and the fact that amplicon sequencing cannot account for marker gene copy number variations (Firkins and Yu 2015, Poretsky et al. 2014). Another critical step in metataxonomics is the construction and maintenance of updated and representative databases to aid accurate taxonomical assignments of rumen phylotypes. Commonly used databases in rumen metataxonomic studies are the Greengenes (DeSantis et al. 2006) and SILVA (Quast et al. 2013) databases for bacteria, RIM-DB (Seedorf et al. 2014) for methanogens, the ureC database (Jin et al. 2017) for ureolytic bacteria, and AF-RefSeq (Paul et al. 2018) for anaerobic fungi.

In such analysis, phylogenetic gene investigations are based on the sequence similarity or, more precisely, on evolutionary divergence between defined taxonomic units (Denman et al. 2018). In earlier studies, a value of 97% sequence similarity was used to define a species level rank at least for the full-length 16S rRNA gene (Stackebrandt and Goebel 1994). With the rise of shorter sequences generated from next-generation sequencing technologies, the previously recommended threshold of 97% has now been revised to a suggested value of 99 to 100% sequence similarities to define the operational taxonomic unit (OTU) (or more commonly now, the Amplicon Sequence Variant) (Callahan et al. 2016). However, using a value of 100% is also likely to generate multiple OTUs from the same species, as a single species (or different strains) may possess multiple copies of an identical 16S rRNA gene (Větrovský and Baldrian 2013). The choice of a suitable variable region along with the similarity threshold of 99% to allow for possible polymorphism effects may be a desirable step to mitigate those limitations (Denman et al. 2018). The most popular variable region currently being targeted to cover both bacterial and archaeal populations is the V4 region (Kozich et al. 2013), but other researchers have

suggested that the V1-V3 and V6-V8 regions should be the choice for rumen bacteria and archaea, respectively (Peter H. Janssen and Marek Kirs 2008, Fuyong Li et al. 2016). While metataxonomics provides insight into the composition of the rumen microbiota, this technique cannot assess the microbial function within a given environment, which needs to be performed using other methods as discussed below. Regardless of this limitation, researchers can still predict the rumen metabolic functions from phylogenetic data through programs derived from the PICRUSt tool (Langille et al. 2013) such as CowPi (Wilkinson et al. 2018) although such tools are still subject to the inherent biases and limitations of amplicon sequencing.

1.4.2 Metagenomics

The concept of metagenomics was first used by Handelsman et al. (1998) to explore the biosynthetic machinery of soil microbiota, and later it became the recommended approach to characterize the potential function of the microbiota directly from their genomes (Marchesi and Ravel 2015). Determining the functional capacity of the rumen microbiota is feasible and can be achieved through random sequencing of all genetic material contained in a sample (metagenomic shotgun sequencing), with the aim of cataloging genes and species (Denman et al. 2018). Metagenomic shotgun sequencing can thus be used to profile the taxonomy, catalogue the functional genes, discover new enzymes and pathways, assemble whole- and fragmented-genomes, and quantify the abundance of functional genomic elements across and between samples (Shinichi Sunagawa et al. 2013, Gupta et al. 2016b, Huws et al. 2018). Metagenomics can provide informative taxonomic and functional profiles using several analytical methods (S. Sunagawa et al. 2013), which include the analysis of informative marker genes (e.g. the 16S rRNA gene) or contigs assembly aligned to databases of reference microbial genomes (Gupta et al. 2016a). Several studies to date have employed those analytical approaches in rumen microbiome investigations, including the first functional metagenomic assessment of the rumen microbiome in pre-ruminant calves (Li et al. 2012) and the effect of feed conversion rate and breeds on the structure and functions of the rumen microbiome (Roehe et al. 2016).

While shotgun metagenomics allows for the study of uncultivable microbial profiles and has become an essential tool for understanding the genomic potential of the rumen microbiome (Morgavi et al. 2013), limitations still apply to this method. Sample collection protocols, DNA extraction techniques and issues around varying levels of genomic DNA for different species can bias the detection of closely related species or strains that can become inadvertently co-assembled by traditional metagenomic approaches (Denman et al. 2018). To overcome these limitations, binning techniques have been developed to exploit the functional capacity of the rumen microbiota through the construction of complete or near complete microbial genomes directly from metagenomic sequencing data (Tyson et al. 2004). The ability to bin genomes from metagenomes stems from an *in silico* approach whereby metagenomic assembled contigs are placed in common bins based on the frequency and abundance of nucleotides, and coverage depth within the sample using programs like PhyloPythiaS, GroopM, and MetaBat (Patil et al. 2011, Kang et al. 2015, Imelfort et al. 2014). Completeness and contamination of metagenomic assembled genomes (MAGs) can then be assessed based on the presence of multiple lineage-specific single-copy marker genes using programs like CheckM or PhyloSift (Darling et al. 2014, Parks et al. 2015). These methodologies also allow for assigning taxonomic information to MAGs from single-copy markers genes concatenated in genome-based taxonomy trees (Denman et al. 2018). The utility of MAGs was demonstrated by Svartström et al. (2017), who assembled 99 microbial genomes from the moose rumen, and by Stewart et al. (2018), who assembled 913 microbial genomes from the rumen of cattle.

In addition to the applications discussed previously, metagenomics can be used to identify novel enzymes in the rumen. The discovery of enzymes in metagenomes can be accomplished using two different strategies: a) sequence-based metagenomic approaches that look for homologous enzymes in metagenomic datasets, and b) functional metagenomic approaches, in which metagenomic libraries are constructed, and clones are screened for enzyme activity (Lam et al. 2015, Sabree et al. 2009). Sequence-based metagenomic techniques rely on the search for genes that code for a particular enzyme through the design of PCR primers or hybridization probes on conserved regions and motifs of known protein families (Ferrer et al. 2009). In contrast,

the functional approach is not dependent on previous genomic knowledge and allows for the discovery of novel enzymes with unexpected peptide sequences (exhibiting classical or new activities) that would not be predicted based on DNA sequencing alone (Distaso et al. 2017).

The diversity of hydrolytic enzymes in a metagenomic library was screened for hydrolase activity for the first time in 2005 (Ferrer et al. 2005), when a total of 22 clones with hydrolytic activities were identified and characterized in the rumen of a cow. Since then, numerous metagenomic studies have reported the diversity of fibrolytic enzymes in ruminants fed forage diets (Brulc et al. 2009, Hess et al. 2011, Wang et al. 2013). While metagenomic techniques have successfully uncovered genes for lignocellulose breakdown, this method cannot provide further information about transcripts that are actively transcribed during lignocellulose breakdown (Rosnow et al. 2017). As described above, shotgun metagenomics can be used for taxonomic identification, functional characterization and discovery of novel enzymes in the rumen microbiome, but it tends to be biased towards the most abundant genes encoded by the most numerically abundant microbial species, and this outcome may not necessarily reflect the importance of a gene, species or strains residing in the metagenome.

1.4.3 Metatranscriptomics

Metatranscriptomics refers to the analysis of expressed RNAs by high-throughput sequencing of the corresponding cDNAs, and is the method used to provide information on the regulation and expression profiles of complex microbiomes (Marchesi and Ravel 2015). While metagenomic shotgun sequencing and MAGs have been advantageous over metataxonomic techniques, they cannot distinguish whether the genomic content of the microbiota is from viable cells as opposed to active function revealed by RNA-based technologies. Currently, metatranscriptomics is considered a reliable approach to identify metabolically active microbial communities (Franzosa et al. 2014) and to find new functions and/or enzymes that would not be otherwise revealed by metagenomic approaches (Rosnow et al. 2017). Metatranscriptome profiling has become popular in rumen microbiology studies in recent years and was initially used to reveal the snapshot of the composition and relative abundance of active species (bacteria and fungi) involved

in the lignocellulose breakdown (Meng Qi et al. 2011, Dai et al. 2015). In addition to taxonomic identification, metatranscriptomic techniques have been used to explore the functional capacity of the rumen microbiome to degrade lignocellulose as exemplified by the investigations of novel CAZymes (e.g., glycoside hydrolases, carbohydrate-binding modules, carbohydrate esterases) involved in the ruminal digestion of recalcitrant diets in ruminants (Fuyong Li and Le Luo Guan 2017, Comtet-Marre et al. 2017, Comtet-Marre et al. 2018).

Currently, most metatranscriptomic analyses employ techniques to deplete the rRNA sequences with the aim of increasing the number of non-rRNA reads in the datasets, as concentrations of rRNA are not consistently correlated with microbial growth and can differ between closely related taxa (Blazewicz et al. 2013). Another reason for rRNA depletion is that rRNA transcripts are highly abundant (~90% of all the read data) and this may bring issues to the analysis of mRNA sequences (Fuyong Li and Le Luo Guan 2017). Thus, metatranscriptomic analysis requires either computational binning of the rRNA genes to identify and remove ribosomal sequences (using total RNA- or mRNA-based methods) or use of kits (probes) to deplete rRNA pre-sequencing (Huws et al. 2018). In the first alternative, computational programs such as SortMeRNA (Kopylova et al. 2012) can be used to sort the filtered metatranscriptomic reads (total RNA-based methods) into fragments of 16S rRNA for taxonomic identification using pipelines like Mothur (Schloss et al. 2009, Fuyong Li et al. 2016) or to obtain mRNA sequences for microbial classification using software like Kraken (Neves et al. 2017, Wood and Salzberg 2014). The second alternative is the use of kits to remove prokaryotic and eukaryotic rRNAs, but a large variety of kits are required to remove rRNAs from complex microbiomes such as the rumen, and this may be costly and laborious (Huws et al. 2018). Additionally, the time required to perform the removal of ribosomal sequences using the kits may cause partial RNA degradation and thus introduce biases in the downstream analysis (Huws et al. 2018). Despite these limitations, bespoke kits used to remove rumen-derived microbial rRNAs were described in a study on dairy cows in France where 18 newly ribosomal depletion probes were designed to remove rumen microbial rRNAs and covered a significant proportion of the rumen bacterial, archaeal, fungal, and protozoal populations enriched for non-rRNA reads (Comtet-Marre et al. 2017). More recently, Li et al. (2019a)

suggested that mRNA-enriched metatranscriptomics should be used for the study of specific genes and/or metabolic pathways with low expression levels, while total RNA-based metatranscriptomics are best for linking compositional and functional profiles of the rumen microbiota to host phenotypes.

While metatranscriptomic analysis is a powerful tool to identify which organisms are present and genes that are expressed in a sample, one major challenge is the *de novo* assembly of the sequencing data to obtain a more reliable annotation of the expressed genetic content of the metatranscriptome (Davids et al. 2016). In addition to the challenges associated with identifying unique transcripts through assembly, metatranscriptomic analysis does not account for variations in the translation and turnover rates of transcripts transcribed into proteins, and thus functional activities captured by metatranscriptomics are poorly inferred from correlations between the gene content (enzyme abundance) and the transcript expression (transcript abundance) (Rosnow et al. 2017). Thus further investigations in protein detection could offer additional information about the microbial activity and protein expression in the rumen.

1.4.4 Metaproteomics

Metaproteomics is a method used to characterize the entire protein content of environmental samples at a given point in time and was introduced by Wilmes and Bond (2004). The method indiscriminately identifies proteins from the environmental samples (metagenomes) and is performed using liquid-chromatography-based separation techniques coupled to mass spectrometry (Marchesi and Ravel 2015). Inference of proteins from the identified peptides and the determination of the taxonomic origin and function of these proteins can then be achieved using protein alignment tools such as UniPept (Mesuere et al. 2012). This method has been successfully used to detect proteins and to identify physiological responses to changes observed in various environmental conditions such as in soil sediments (Benndorf et al. 2009), rhizosphere (Wu et al. 2011) and human distal gut microbiota (Verberkmoes et al. 2008). However, a limited number of metaproteomic studies have been published for the rumen. The technical issues that have been presented to explain the shortcomings of metaproteomics for predicting the function of rumen proteins include the interference of polyphenolic compounds of the diets

with protein isolation procedures. Phenolic compounds (e.g., tannins and humic acids) co-precipitate with proteins and alter gel mobility, resulting in unresolved smears rather than distinct protein bands, and thus limit the resolution of the peptides from the rumen microbiome (Snelling and Wallace 2017).

Despite these limitations, Snelling and Wallace (2017) identified rumen microbial proteins, such as actin, alpha and beta tubulins, and axonemal isoforms dynein light chain, which are involved in the locomotion of ciliate protozoa. Removal of protozoa from digesta before protein extraction revealed the prokaryotic metaproteome and the results showed a predominance of enzymes of the central metabolism originating from the Firmicutes and Bacteroidetes phyla (Snelling and Wallace 2017). Another possibility is to combine shotgun-metaproteomics with 16S rRNA amplicon-based methods to unveil the metaproteome expressed by the digesta-associated microbiota in cattle undergoing dietary changes. By using this approach, Deusch et al. (2017) identified over 8000 bacterial and 350 archaeal proteins in the rumen, further improving our understanding of the complicated interplay among rumen microbes, proteins and their adaptation to various fermentation substrates (e.g., starch, cellulose, hemicellulose). More rumen microbiome research is likely to be published in the upcoming years as bioinformatic and technical progress enhance the metaproteomic coverage and maximize the focus on microbiome-derived peptides.

1.4.5 Future perspectives in omics technologies to study the rumen microbiome

Each of the previously described technological approaches are powerful tools in assessing the microbial composition and functional potential of the rumen microbiome, but our understanding could be further enhanced if the information generated by those technologies were integrated into multivariate models (Huws et al. 2018). However, this integration is a complicated process owing to the complex and heterogeneous nature of the available datasets generated by a wide variety of omics approaches (Fan et al. 2014). Although challenging, there are instances of success in the literature showing the benefits of integrating a varied array of data types generated from omics technologies in rumen microbiome studies. Hart et al. (2018) compared metaproteomic data with published genomic datasets of the rumen microbiome and improved protein identification that could

be impossible without integrating the information generated by the different datasets. Advances in statistical methods could also offer more opportunities to integrate large-scale molecular omics datasets and assess microbial interactions at multiple functional levels.

Despite these developments, there is still a need to increase the number of representative isolates in the microbial culture collections of rumen origin. The "Hungate1000" project recently sequenced 420 representatives of rumen microbes, and thus provided significant input of data for the scientific community (Seshadri et al. 2018). Notwithstanding all of these efforts, it has been estimated that only 3.6% of the OTUs (61 out of 1,698 OTUs) of the sequenced rumen samples have representative isolates in the Ribosomal Database project and only 117 bacterial species (not including different strains) of rumen origin are available from international culture collections (Nordberg et al. 2014, Seshadri et al. 2018, Zehavi et al. 2018, Cole et al. 2014). Thus, it is imperative the adoption of complementary approaches (e.g., single-cell genomics) to sequence the DNA of the whole microbiota in order to study the genomes of as-yet uncultured species from the rumen. Single-cell genomics analysis aims to characterize the genomic variability among individual cells, and thus it could be used to reconstruct cellular ancestries in the form of a lineage tree (Shapiro et al. 2013). In addition to single-cell genomics technologies, metabolomics could be used to profile (qualitatively and quantitatively) the metabolites of a microbial community to prospect novel compounds and metabolic pathways in the rumen microbiome (Yi et al. 2016, Deusch et al. 2017, Deusch et al. 2015). These technologies are in constant development, and the continued technical and analytical advances in the field of molecular biology are likely to cause the emergence of new technologies in the coming years.

1.4.6 Future perspectives in molecular techniques to investigate lignocellulolytic enzymes

While metagenomics and metatranscriptomics have been successful to identify novel enzymes responsible for lignocellulose degradation, these techniques have limitations to evaluate changes in protein abundance, isoform expression, turnover rates, and post-translational modifications and interactions (Larance and Lamond 2015). These

limitations have been overcome by high-throughput screening methods developed in recent years, and with the progress in robotics and functional assays. In discovery-based studies, quantitative MS and MS/MS measurements and nuclear magnetic resonance spectroscopy are the most popular methods to identify enzymes and metabolites involved in lignocellulose degradation (Rosnow et al. 2017). Efficient screening tools such as fluorescence-activated cell sorting in combination with substrate-induced gene expression have been developed to perform direct measurements of active enzyme in a sample and to identify enzyme functions necessary for lignocellulose breakdown (Distaso et al. 2017). These tools are still in their infancy, and as the new technical and analytical advancements are achieved, we will see a continued breakthrough in our understanding of how microbial enzymes accomplish lignocellulose breakdown.

1.5 Statistical challenges associated with the studies of the rumen microbiome

Datasets generated using omics technologies discussed in previous sections are inherently compositional, a feature which is known to be problematic and should not be ignored by data analysts (Pawlowsky‐Glahn et al. 2015, Aitchison 1982, Fernandes et al. 2013, Warton et al. 2012, Gloor et al. 2017, Pearson 1897). Compositional data is a type of dataset comprised in a mathematical space known as simplex space, where the features (OTUs, genes, etc.) in each sample hold proportions of a unit varying between 0 and 1 (Aitchison 1982). Unlike the simplex space, the Euclidean space does not exhibit constraints between 0 and 1, but can accept any real number along its dimensions (Fernandes et al. 2014). Thus, the analysis of microbiome data requires statistical methods accounting for the simplex structure of compositional datasets and excludes standard statistical techniques (including Pearson correlations, Principal Component Analysis, linear regression, etc.) that use the assumptions of the Euclidian space (Pearson 1897, Aitchison 1982, Lovell et al. 2015, Gloor and Reid 2016, Fernandes et al. 2014). Despite these limitations, those traditional statistical methods are still commonly used by the scientific community.

The original problem in analyzing compositional data was first identified by Pearson in 1897 (Pearson 1897) when he realized that the count values per feature in a compositional data are not independent, with the value of one feature necessarily

restricting the value of at least one other feature (Fernandes et al. 2014). This property can lead to negative correlation biases and false univariate inferences in compositional data, which renders invalid any correlation- or covariance-based results (Pearson 1897, Aitchison 1982, Fernandes et al. 2013). An easy analogy to explain this distortion is the "see-saw effect", in which a change in the abundance of one feature results in a biased correlation between the other features (one goes up, another goes down). In addition to the possibility of obtaining spurious results, investigators should acknowledge that the relationship between absolute abundance in the environment and the relative abundance after sequencing is not equivalent in compositional datasets because the number of reads obtained for a sample is determined by the capacity of the instrument and not by the actual number of molecules of DNA in the environment (Mandal et al. 2015, Gloor and Reid 2016). Therefore, compositional datasets are very different from datasets composed of ordinary numbers that can take any value, and treating high-throughput sequencing data as compositional is rather intuitive if one considers that the number of counts in such datasets reflects the proportion of counts per feature per sample multiplied by the sequencing depth (Fernandes et al. 2014, Gloor et al. 2017).

Sequencing depth (the total number of counts observed) between samples is another significant confounder of the analysis, as abundance issues arise around the variation in the number of sequences obtained for each sample. Rarefying or subsampling the read counts of each sample to a defined level across samples excludes lower abundant features (OTUs, genes, etc.) leading to a loss of precision (McMurdie and Holmes 2014). If the researcher chooses to use the entire dataset (without rarefying), s/he must account for the magnitude of sequence depth between samples and usually needs to employ a transformation or scaling method (e.g., trimmed mean of M values – TMM and the median methods) (McMurdie and Holmes 2014, Weiss et al. 2017, Robinson et al. 2010, Anders and Huber 2010, Robinson and Oshlack 2010, Lovell et al. 2015, Love et al. 2014). Methods for the identification of differentially abundant OTUs associated with a given phenotype or treatment should not use models that apply Poisson distribution because it is too restrictive to deal with overdispersion (Anders and Huber 2010). To address the overdispersion problem, researchers have proposed the use of negative binomial distributions, but this tends to increase the false discovery rate arising

from the compositional nature of the data (Lovell et al. 2015, Anders and Huber 2010, Gloor et al. 2017). Taken altogether, the data analyst should be careful while analyzing microbiome data as it exhibits a compositional structure that needs to be dealt with appropriately in the statistical analysis.

1.5.1 Alternative techniques to study microbiome data

As outlined above, data collected from high-throughput sequencing platforms present challenges to ecological and statistical analysis, and to circumvent these issues, alternative statistical methods have been developed to substitute the standard statistical approaches in the analysis of compositional data. The reader is directed to a review by Gloor et al. (2017) who provided a more in-depth discussion of the application of these techniques in the analysis of microbiome data as alternatives to the standard statistical approaches. These alternative methods include initial normalization of the count data using log-ratio transformations (centered or isometric) rather than rarefaction. Another significant change in the analysis steps is the replacement of beta diversity analysis using Aitchison calculations of distances for Bray Curtis. Researchers have also suggested substituting *philr* transform for the unifrac distance metric when analyzing phylogenetic trees (Silverman et al. 2017), and that beta diversity variance is visualized based on compositional principal component biplot rather than principal co-ordinate plots (Gloor et al. 2017). Correlation analysis to assess the extent to which a pair of random variables are proportional should be performed with appropriate metrics such as the “goodness-of-fit to proportionality” statistic ϕ rather than Pearson or Spearman correlations (Lovell et al. 2015). Finally, identification of differentially abundant features and microbial signatures has been advised using ANCOM (Mandal et al. 2015) and *MixMC* (Cao et al. 2016). These two methods will be briefly discussed below.

1.5.1.1 ANCOM

The Analysis of Composition of Microbiomes (ANCOM) (Mandal et al. 2015) is a statistical procedure that compares the Aitchison’s log-ratio of the abundance of each taxon with the abundance of all remaining taxa one at a time. Then, differential tests (e.g., Mann-Whitney *U*, ANOVA, ANOVA with Linear Mixed Effect Models, Friedman, Kruskal-Wallis, Wilcoxon tests) are calculated on each log ratio to reveal differences in the relative

abundance of a taxon between two ecosystems. If there are “m” taxa, then for each taxon ANCOM performs “m-1” tests and the final significance of each test for a taxon is determined using Benjamini-Hochberg (Benjamini and Hochberg 1995) algorithm to control for false discovery rates. For each taxon, ANCOM counts the number of tests among the m-1 tests and obtains a count random variable *W* that represents the number of null among the m-1 tests that are rejected. To deal with the sparsity of the data, ANCOM uses an arbitrary pseudo count value of 0.001 to replace the zero counts and calculate the log-ratios. For drawing inferences regarding taxon abundance in the ecosystem, ANCOM has been suggested as a reliable method to control the identification of false positives and has been recently incorporated into the QIIME2 pipeline (Caporaso et al. 2010) (https://qiime2.org). ANCOM was recently implemented in a bioinformatic pipeline developed by our group (Neves et al. 2017) to detect differentially abundant taxa identified by Kraken (Wood and Salzberg 2014) and Mothur (Schloss et al. 2009) in the rumen metatranscriptome.

1.5.1.2 MixMC

MixMC (Multivariate Statistical Framework to Gain Insight into Microbial Communities) (Cao et al. 2016) is a framework that takes into account the inherent characteristics of microbiome data (sparsity and compositionality) to identify microbial signatures associated with their environment, and it is currently implemented in the package *mixOmics* (Rohart et al. 2017). In *MixMC*, the method Sparse Partial Least Square Discriminant Analysis (sPLS-DA) is associated with centered log-ratio (CLR) transformations to project the data from a simplex space to a Euclidian space and includes a multilevel decomposition for repeated measurements designs that are commonly encountered in microbiome studies (Cao et al. 2016). To account for subject variability, the data variance is decomposed into *within variation* (due to habitat) and *between-subject variation*. This is an appropriate analytical step towards detecting subtle differences between samples when high inter-subject variability is present due to sampling repeatedly performed on the same subjects and in multiple habitats (Westerhuis et al. 2010, Liquet et al. 2012, Cao et al. 2016). Before the datasets are log-transformed (CLR) and analyzed by the sPLS-DA models, preprocessing and normalization (e.g., total

sum scaling) steps are performed to account for the sparsity of the dataset and uneven sequencing depths across samples.

1.5.2 Current challenges when comparing results across studies

While next-generation sequencing resulted in an explosion of publications exploring the rumen microbial diversity and functions in the last decades, interpretation of the data generated across multiple studies are still hampered by the lack of standardization in the bioinformatic and statistical procedures employed by the different research groups. As discussed previously, differences among studies exist for DNA extraction methods, primers, PCR cycling parameters, and downstream bioinformatic analysis which make comparisons across studies problematic and impractical at the moment. One instance of this problem appeared when the rumen microbiome of efficient cattle was compared across studies in order to find consensus microbial genes that could serve as global biomarkers for predicting ruminant feed efficiency and methane emissions. Huws et al. (2018) reported that microbial gene correlations with RFI described by Shabat et al. (2016) overlapped with those of Fuyong Li and Le Luo Guan (2017) only in relation to a lower abundance of genes involved in amino acid metabolism in the rumen of feed efficient animals. However, genes related to methanogenesis did not show a consensus between the datasets of Shabat et al. (2016) and Fuyong Li and Le Luo Guan (2017), indicating that a standardization in the analysis is needed to compare across studies and promote reproducibility of the results.

Considering the limitations of the omics technologies and the lack of awareness to use robust statistical tools to analyze compositional data, there is an urgent need for guidelines and practices that standardize rumen microbiome studies from the wet laboratory to the publication stage. Aspects of the analysis that are important to standardize include the methods of OTU picking, the databases (public or customized) and algorithms for taxonomy classification, cutoffs for taxa inclusion/exclusion, and the statistical methods used to analyze the microbiome data (Goodrich et al. 2017). Irrespective, only after the standardization of the workflow (from sample collection to the analysis) is completed, there will be more reliable comparisons of results across studies.

1.6. Knowledge gaps, hypothesis and objectives

The significant advances in omics technologies in recent years has revolutionized our understanding of the rumen microbiome, its role in feed efficiency, and the degree to which it is influenced by the host genetic background. However, despite these steps forward, there remain substantial gaps in our knowledge. The broad aim of this thesis was to fill the gaps concerning the driving forces that influence the relationship between the rumen microbiota and host individual variation, and how their interactive effects on animal productivity contribute to the identification of cattle with improved feed efficiency. Moreover, this thesis fills the knowledge gaps concerning the impact of mRNA-based metatranscriptomics on the analysis of rumen taxonomic profiles. Further assessing such methodologies will contribute to elucidate the link between these taxonomic profiles and feed efficiency at the RNA level. Another gap in the literature is the lack of a strategy for the discovery of lignocellulosic enzymes through the targeted functional profiling of CAZyme families, as the discovery of these biocatalysts is usually performed using metagenomic screening or metatranscriptomic analysis that do not prioritize the identification of novel enzymes based on their ecological relevance in the microbiome. Therefore, further investigations of the microbial and functional dynamics of the rumen metatranscriptome are needed, and examining how they relate to the ruminant ability to degrade lignocellulosic biomass will be critical in designing innovative strategies to discover unknown enzymes for the breakdown, biosynthesis or modification of lignocellulosic biomass.

The overall hypotheses for this thesis were that the stratification of dietary responses obtained from the magnitude of change in baseline rumen microbiota can be used to identify feed efficient cattle, and mRNA-based metatranscriptomic methods can be applied to characterize the microbial composition, diversity and functional profiles of the rumen microbiome in cattle with different feed efficiencies. Moreover, I hypothesized that feed efficiency affects the structure of the rumen microbiota and subsequently the expression of enzymes associated with lignocellulose degradation. The research presented in this thesis contributes to our understanding of the dynamics of the rumen microbiome in cattle undergoing dietary changes and also sheds light on the factors within

the microbiome and the host that maximize the degradation of lignocellulosic biomass. The objectives of this thesis were as follows:

1. To investigate the dynamics of rumen microorganisms in cattle raised under different feeding regimens and understand the relationship among the abundance of these microorganisms, host individuality and phenotypic traits;

2. To compare and contrast methodologies (mRNA- vs. total rRNA-based methods) to assess the taxonomic profiles of the rumen microbiota and to investigate the impact of the comparative analysis of both analytical approaches on the rumen microbial classification obtained from cattle exhibiting different feed efficiencies;

3. To characterize active microbial functional signatures differentiating breeds of beef cattle, and identify specific taxonomic microbial groups and functions associated with feed efficiency;

4. To identify and validate biological associations (composition and functions) between the rumen metatranscriptome and feed efficiency in beef cattle, and use this knowledge to discover new enzymes for lignocellulose degradation.

1.7 Literature Cited

Aitchison, J. (1982). The statistical-analysis of compositional data. *Journal of the Royal Statistical Society Series B-Methodological* 44(2)**,** 139-177.

Akin, D.E., Borneman, W.S., and Windham, W.R. (1988). Rumen fungi - morphological types from georgia cattle and the attack on forage cell-walls. *Biosystems* 21(3-4)**,** 385-391. doi: 10.1016/0303-2647(88)90037-8.

Alemu, A.W., Vyas, D., Manafiazar, G., Basarab, J.A., and Beauchemin, K.A. (2017). Enteric methane emissions from low– and high–residual feed intake beef heifers measured using GreenFeed and respiration chamber techniques12. *Journal of Animal Science* 95**,** 3727-3737. doi: 10.2527/jas.2017.1501.

Alexandratos, N., and Bruinsma, J. (2012). World Agriculture Towards 2030/2050: The 2012 Revision. *FAO - Food and Agriculture Organization of the United Nations*.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology* 11(10)**,** R106. doi: 10.1186/gb-2010-11-10-r106.

AOAC (2000). *Official Methods of Analysis (17th).* Gaithersburg, MD: Association of Official Analytical Chemists.

Archer, J.A., Richardson, E.C., Herd, R.M., and Arthur, P.F. (1999). Potential for selection to improve efficiency of feed use in beef cattle: a review. *Australian Journal of Agricultural Research* 50(2)**,** 147-162. doi: https://doi.org/10.1071/A98075.

Artzi, L., Bayer, E.A., and Moraïs, S. (2016). Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nature Reviews Microbiology* 15**,** 83. doi: 10.1038/nrmicro.2016.164.

Bader, J., Mast-Gerlach, E., Popovic, M.K., Bajpai, R., and Stahl, U. (2010). Relevance of microbial coculture fermentations in biotechnology. *Journal of Applied Microbiology* 109(2)**,** 371-387. doi: 10.1111/j.1365-2672.2009.04659.x.

Bae, H.D., McAllister, T.A., Kokko, E.G., Leggett, F.L., Yanke, L.J., Jakober, K.D., et al. (1997). Effect of silica on the colonization of rice straw by ruminal bacteria. *Animal Feed Science and Technology* 65(1)**,** 165-181. doi: https://doi.org/10.1016/S0377-8401(96)01093-0.

Bashiardes, S., Godneva, A., Elinav, E., and Segal, E. (2018). Towards utilization of the human genome and microbiome for personalized nutrition. *Current Opinion in Biotechnology* 51**,** 57-63. doi: 10.1016/j.copbio.2017.11.013.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *2015* 67(1)**,** 48. doi: 10.18637/jss.v067.i01.

Bayer, E.A., Chanzy, H., Lamed, R., and Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology* 8(5)**,** 548-557. doi: https://doi.org/10.1016/S0959-440X(98)80143-7.

Bayley, D.P., Correia, J.D., Jarrell, K.F., and Thomas, N.A. (1999). Recent Excitement about the Archaea: The Archaea are valuable for studying basic biological questions and have novel biotechnology applications. *BioScience* 49(7)**,** 530-541. doi: 10.2307/1313474.

Beg, Q., Kapoor, M., Mahajan, L., and Hoondal, G. (2001). Microbial xylanases and their industrial applications: a review. *Applied Microbiology and Biotechnology* 56(3)**,** 326-338. doi: 10.1007/s002530100704.

Belanche, A., Abecia, L., Holtrop, G., Guada, J.A., Castrillo, C., de la Fuente, G., et al. (2011). Study of the effect of presence or absence of protozoa on rumen fermentation and microbial protein contribution to the chyme1. *Journal of Animal Science* 89(12)**,** 4163-4174. doi: 10.2527/jas.2010-3703.

Belanche, A., Doreau, M., Edwards, J.E., Moorby, J.M., Pinloche, E., and Newbold, C.J. (2012). Shifts in the Rumen Microbiota Due to the Type of Carbohydrate and Level of Protein Ingested by Dairy Cattle Are Associated with Changes in Rumen Fermentation. *Journal of Nutrition* 142(9)**,** 1684-1692. doi: 10.3945/jn.112.159574.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* 57(1)**,** 289-300.

Benndorf, D., Vogt, C., Jehmlich, N., Schmidt, Y., Thomas, H., Woffendin, G., et al. (2009). Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments. *Biodegradation* 20(6)**,** 737-750. doi: 10.1007/s10532-009-9261-3.

Berry, D.P., and Crowley, J.J. (2013). CELL BIOLOGY SYMPOSIUM: Genetics of feed efficiency in dairy and beef cattle1. *Journal of Animal Science* 91(4)**,** 1594-1613. doi: 10.2527/jas.2012-5862.

Bhat, M.K., and Bhat, S. (1997). Cellulose degrading enzymes and their potential industrial applications. *Biotechnology Advances* 15(3)**,** 583-620. doi: https://doi.org/10.1016/S0734-9750(97)00006-2.

Bhatia, L., Johri, S., and Ahmad, R. (2012). An economic and ecological perspective of ethanol production from renewable agro waste: a review. *Amb Express* 2. doi: 10.1186/2191-0855-2-65.

Blazewicz, S.J., Barnard, R.L., Daly, R.A., and Firestone, M.K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The Isme Journal* 7**,** 2061. doi: 10.1038/ismej.2013.102.

Borneman, W.S., Hartley, R.D., Morrison, W.H., Akin, D.E., and Ljungdahl, L.G. (1990). Feruloyl and p-coumaroyl esterase from anaerobic fungi in relation to plant cell wall degradation. *Applied Microbiology and Biotechnology* 33(3)**,** 345-351. doi: 10.1007/BF00164534.

Brooker, J.D., O'Donovan, L.A., Skene, I., Clarke, K., Blackall, L., and Muslera, P. (1994). Streptococcus caprinus sp.nov., a tannin-resistant ruminal bacterium from feral goats. *Letters in Applied Microbiology* 18(6)**,** 313-318. doi: 10.1111/j.1472-765X.1994.tb00877.x.

Brulc, J.M., Antonopoulos, D.A., Berg Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., et al. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences* 106(6)**,** 1948-1953. doi: 10.1073/pnas.0806191105.

Bryant, M.P., and Burkey, L.A. (1953). Cultural Methods and Some Characteristics of Some of the More Numerous Groups of Bacteria in the Bovine Rumen. *Journal of Dairy Science* 36(3)**,** 205-217. doi: https://doi.org/10.3168/jds.S0022-0302(53)91482-9.

Bryant, M.P., and Robinson, I.M. (1961). An Improved Nonselective Culture Medium for Ruminal Bacteria and Its Use in Determining Diurnal Variation in Numbers of Bacteria in the Rumen. *Journal of Dairy Science* 44(8)**,** 1446-1456. doi: https://doi.org/10.3168/jds.S0022-0302(61)89906-2.

Burnham, K.P., Anderson, D.R., and Burnham, K.P. (2002). *Model selection and multimodel inference : a practical information-theoretic approach.* New York : Springer, c2002.

2nd ed.

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

https://www.nature.com/articles/nmeth.3869 - supplementary-information.

Cammack, K.M., Austin, K.J., Lamberson, W.R., Conant, G.C., and Cunningham, H.C. (2018). Tiny but mighty: The role of the rumen microbes in livestock production. *J Anim Sci*. doi: 10.1093/jas/skx053.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* 37**,** D233-D238. doi: 10.1093/nar/gkn663.

Cao, K.A.L., Costello, M.E., Lakis, V.A., Bartolo, F., Chua, X.Y., Brazeilles, R., et al. (2016). MixMC: A Multivariate Statistical Framework to Gain Insight into Microbial Communities. *Plos One* 11(8). doi: 10.1371/journal.pone.0160169.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7(5)**,** 335-336. doi: 10.1038/nmeth.f.303.

Carberry, C.A., Kenny, D.A., Han, S., McCabe, M.S., and Waters, S.M. (2012a). Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. *Appl Environ Microbiol* 78(14)**,** 4949-4958. doi: 10.1128/aem.07759-11.

Carberry, C.A., Kenny, D.A., Kelly, A.K., and Waters, S.M. (2014a). Quantitative analysis of ruminal methanogenic microbial populations in beef cattle divergent in phenotypic residual feed intake (RFI) offered contrasting diets. *J Anim Sci Biotechnol* 5(1)**,** 41. doi: 10.1186/2049-1891-5-41.

Carberry, C.A., Kenny, D.A., Sukkyan, H., McCabe, M.S., and Waters, S.M. (2012b). Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. *Applied and Environmental Microbiology* 78(14)**,** 4949-4958.

Carberry, C.A., Waters, S.M., Kenny, D.A., and Creevey, C.J. (2014b). Rumen methanogenic genotypes differ in abundance according to host residual feed intake phenotype and diet type. *Appl Environ Microbiol* 80(2)**,** 586-594. doi: 10.1128/aem.03131-13.

Chafe, S.C. (1969). The fine structure of the collenchyma cell wall. *Planta* 90(1)**,** 12-21. doi: 10.1007/BF00389292.

Cheng, K.J., McCowan, R.P., and Costerton, J.W. (1979). Adherent epithelial bacteria in ruminants and their roles in digestive tract function. *Am J Clin Nutr* 32(1)**,** 139-148. doi: 10.1093/ajcn/32.1.139.

Ciucci, S., Ge, Y., Durán, C., Palladini, A., Martínez-Sánchez, L.M., Cannistraci, C.V., et al. (2017). Enlightening discriminative network functional modules behind principal component analysis separation in differential-omic science studies. *Scientific Reports* 7. doi: 10.1038/srep43946.

Clark, J.H., Budarin, V., Deswarte, F.E.I., Hardy, J.J.E., Kerton, F.M., Hunt, A.J., et al. (2006). Green chemistry and the biorefinery: a partnership for a sustainable future. *Green Chemistry* 8(10)**,** 853-860. doi: 10.1039/B604483M.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., et al. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research* 42(D1)**,** D633-D642. doi: 10.1093/nar/gkt1244.

Comtet-Marre, S., Chaucheyras-Durand, F., Bouzid, O., Mosoni, P., Bayat, A.R., Peyret, P., et al. (2018). FibroChip, a Functional DNA Microarray to Monitor Cellulolytic and Hemicellulolytic Activities of Rumen Microbiota. *Frontiers in Microbiology* 9(215). doi: 10.3389/fmicb.2018.00215.

Comtet-Marre, S., Parisot, N., Lepercq, P., Chaucheyras-Durand, F., Mosoni, P., Peyretaillade, E., et al. (2017). Metatranscriptomics Reveals the Active Bacterial and Eukaryotic Fibrolytic Communities in the Rumen of Dairy Cow Fed a Mixed Diet. *Frontiers in Microbiology* 8(67). doi: 10.3389/fmicb.2017.00067.

Creevey, C.J., Kelly, W.J., Henderson, G., and Leahy, S.C. (2014). Determining the culturability of the rumen bacterial microbiome. *Microbial Biotechnology* 7(5)**,** 467-479. doi: 10.1111/1751-7915.12141.

Crews, D.H., Jr., Shannon, N. H., Genswein, B. M. A., Crews, R. E., Johnson, C. M. and Kendrick, B. A (2005). Genetic parameters for net feed efficiency of beef cattle measured during postweaning growing versus finishing periods. *Proc. West Sec. Am. Soc. Anim. Sci* 54(54)**,** 125-128.

Dai, X., Tian, Y., Li, J., Su, X., Wang, X., Zhao, S., et al. (2015). Metatranscriptomic Analyses of Plant Cell Wall Polysaccharide Degradation by Microorganisms in the Cow Rumen. *Applied and Environmental Microbiology* 81(4)**,** 1375.

Darling, A.E., Jospin, G., Lowe, E., Matsen, F.A.I.V., Bik, H.M., and Eisen, J.A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2**,** e243. doi: 10.7717/peerj.243.

Davids, M., Hugenholtz, F., Martins dos Santos, V., Smidt, H., Kleerebezem, M., and Schaap, P.J. (2016). Functional Profiling of Unfamiliar Microbial Communities Using a Validated De Novo Assembly Metatranscriptome Pipeline. *PLOS ONE* 11(1)**,** e0146423. doi: 10.1371/journal.pone.0146423.

Denman, S.E., Martinez Fernandez, G., Shinkai, T., Mitsumori, M., and McSweeney, C.S. (2015). Metagenomic analysis of the rumen microbial community following inhibition of methane formation by a halogenated methane analog. *Frontiers in Microbiology* 6(1087). doi: 10.3389/fmicb.2015.01087.

Denman, S.E., and McSweeney, C.S. (2006). Development of a real-time PCR assay for monitoring anaerobic fungal and cellulolytic bacterial populations within the rumen. *Fems Microbiology Ecology* 58(3)**,** 572-582. doi: 10.1111/j.1574-6941.2006.00190.x.

Denman, S.E., Morgavi, D.P., and McSweeney, C.S. (2018). Review: The application of omics to rumen microbiota function. *animal* 12(s2)**,** s233-s245. doi: 10.1017/S175173111800229X.

Deppenmeier, U. (2002). The unique biochemistry of methanogenesis. *Prog Nucleic Acid Res Mol Biol* 71**,** 223-283.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72(7)**,** 5069-5072. doi: 10.1128/aem.03006-05.

Deusch, S., Camarinha-Silva, A., Conrad, J., Beifuss, U., Rodehutscord, M., and Seifert, J. (2017). A Structural and Functional Elucidation of the Rumen Microbiome Influenced by Various Diets and Microenvironments. *Frontiers in Microbiology* 8(1605). doi: 10.3389/fmicb.2017.01605.

Deusch, S., Tilocca, B., Camarinha-Silva, A., and Seifert, J. (2015). News in livestock research — use of Omics-technologies to study the microbiota in the gastrointestinal tract of farm animals. *Computational and Structural Biotechnology Journal* 13**,** 55-63. doi: https://doi.org/10.1016/j.csbj.2014.12.005.

Distaso, M.A., Tran, H., Ferrer, M., and Golyshin, P.N. (2017). "Metagenomic Mining of Enzyme Diversity," in *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals,* ed. S.Y. Lee. (Cham: Springer International Publishing), 245-269.

Doi, R.H., and Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nature Reviews Microbiology* 2**,** 541. doi: 10.1038/nrmicro925.

Dollhofer, V., Podmirseg, S.M., Callaghan, T.M., Griffith, G.W., and Fliegerova, K. (2015). Anaerobic Fungi and Their Potential for Biogas Production. *Advances in biochemical engineering/biotechnology* 151**,** 41-61. doi: 10.1007/978-3-319-21993-6_2.

Duffield, T., Plaizier, J.C., Fairfield, A., Bagg, R., Vessie, G., Dick, P., et al. (2004). Comparison of techniques for measurement of rumen pH in lactating dairy cows. *Journal of Dairy Science* 87(1)**,** 59-66.

Edwards, J.E., McEwan, N.R., Travis, A.J., and Wallace, R.J. (2004). 16S rDNA library-based analysis of ruminal bacterial diversity. *Antonie van Leeuwenhoek* (3)**,** 263.

El Enshasy, H.A., Kandiyil, S.K., Malek, R., and Othman, N.Z. (2016). "Microbial Xylanases: Sources, Types, and Their Applications," in *Microbial Enzymes in Bioconversions of Biomass,* ed. V.K. Gupta. (Cham: Springer International Publishing), 151-213.

Fan, J., Han, F., and Liu, H. (2014). Challenges of Big Data analysis. *National Science Review* 1(2)**,** 293-314. doi: 10.1093/nsr/nwt032.

Fernandes, A.D., Macklaim, J.M., Linn, T.G., Reid, G., and Gloor, G.B. (2013). ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLOS ONE* 8(7)**,** e67019. doi: 10.1371/journal.pone.0067019.

Fernandes, A.D., Reid, J.N.S., Macklaim, J.M., McMurrough, T.A., Edgell, D.R., and Gloor, G.B. (2014). Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2(1)**,** 15. doi: 10.1186/2049-2618-2-15.

Fernando, S.C., Purvis, H.T., II, Najar, F.Z., Sukharnikov, L.O., Krehbiel, C.R., Nagaraja, T.G., et al. (2010). Rumen Microbial Population Dynamics during Adaptation to a High-Grain Diet. *Applied and Environmental Microbiology* 76(22)**,** 7482-7490. doi: 10.1128/aem.00388-10.

Ferrer, M., Beloqui, A., Timmis, K.N., and Golyshin, P.N. (2009). Metagenomics for Mining New Genetic Resources of Microbial Communities. *Journal of Molecular Microbiology and Biotechnology* 16(1-2)**,** 109-123. doi: 10.1159/000142898.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A.P.M.D., et al. (2005). Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology* 7(12)**,** 1996-2010. doi: 10.1111/j.1462-2920.2005.00920.x.

Finlay, B.J., Esteban, G., Clarke, K.J., Williams, A.G., Embley, T.M., and Hirt, R.P. (1994). Some rumen ciliates have endosymbiotic methanogens. *FEMS Microbiology Letters* 117(2)**,** 157-161.

Firkins, J.L., and Yu, Z. (2015). RUMINANT NUTRITION SYMPOSIUM: How to use data on the rumen microbiome to improve our understanding of ruminant nutrition. *J Anim Sci* 93(4)**,** 1450-1470. doi: 10.2527/jas.2014-8754.

Ford, C.W., Elliott, R., and Maynard, P.J. (1987). The effect of chlorite delignification on digestibility of some grass forages and on intake and rumen microbial activity in sheep fed barley straw. *Journal of Agricultural Science* 108**,** 129-136.

Fouts, D.E., Szpakowski, S., Purushe, J., Torralba, M., Waterman, R.C., MacNeil, M.D., et al. (2012). Next generation sequencing to define prokaryotic and fungal diversity

in the bovine rumen. *PLoS One* 7(11), e48289. doi: 10.1371/journal.pone.0048289.

Franzosa, E.A., Morgan, X.C., Segata, N., Waldron, L., Reyes, J., Earl, A.M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences* 111(22), E2329-E2338. doi: 10.1073/pnas.1319284111.

Gao, A., Wang, H., Yang, J., and Shi, C. (2013). "The effects of elimination of fungi on microbial population and fiber degradation in sheep rumen," in *Progress in Environmental Protection and Processing of Resource, Pts 1-4,* eds. X. Tang, W. Zhong, D. Zhuang, C. Li & Y. Liu.), 224-231.

Gharechahi, J., and Salekdeh, G.H. (2018). A metagenomic analysis of the camel rumen's microbiome identifies the major microbes responsible for lignocellulose degradation and fermentation. *Biotechnology for Biofuels* 11(1), 216. doi: 10.1186/s13068-018-1214-9.

Gijzen, H.J., Lubberding, H.J., Gerhardus, M.J.T., and Vogels, G.D. (1988). Contribution of rumen protozoa to fibre degradation and cellulase activity in vitro. *FEMS Microbiology Letters* 53(1), 35-43. doi: 10.1111/j.1574-6968.1988.tb02645.x.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology* 8(2224). doi: 10.3389/fmicb.2017.02224.

Gloor, G.B., and Reid, G. (2016). Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Canadian Journal of Microbiology* 62(8), 692-703. doi: 10.1139/cjm-2015-0821.

Goel, G., Puniya, A.K., Aguilar, C.N., and Singh, K. (2005). Interaction of gut microflora with tannins in feeds. *Naturwissenschaften* 92(11), 497-503. doi: 10.1007/s00114-005-0040-7.

Goodrich, J.K., Davenport, E.R., Clark, A.G., and Ley, R.E. (2017). The Relationship Between the Human Genome and Microbiome Comes into View. *Annual Review of Genetics* 51(1), 413-433. doi: 10.1146/annurev-genet-110711-155532.

Gordon, G.L.R., and Phillips, M.W. (1992). Extracellular pectin lyase produced by Neocallimastix sp. LM1, a rumen anaerobic fungus. *Letters in Applied Microbiology* 15(3), 113-115. doi: 10.1111/j.1472-765X.1992.tb00740.x.

Gordon, G.L.R., and Phillips, M.W. (1993). Removal of anaerobic fungi from the rumen of sheep by chemical treatment and the effect on feed consumption and in-vivo fiber digestion. *Letters in Applied Microbiology* 17(5), 220-223. doi: 10.1111/j.1472-765X.1993.tb01451.x.

Gordon, G.L.R., and Phillips, M.W. (1998). The role of anaerobic gut fungi in ruminants. *Nutrition Research Reviews* 11(1), 133-168. doi: 10.1079/nrr19980009.

Greve, L.C., Labavitch, J.M., and Hungate, R.E. (1984). alpha-L-arabinofuranosidase from Ruminococcus albus 8: purification and possible role in hydrolysis of alfalfa cell wall. *Applied and Environmental Microbiology* 47(5), 1135-1140.

Grubb, J.A., and Dehority, B.A. (1975). Effects of an abrupt change in ration from all roughage to high concentrate upon rumen microbial numbers in sheep. *Applied Microbiology* 30(3), 404-412.

Gruninger, R.J., Puniya, A.K., Callaghan, T.M., Edwards, J.E., Youssef, N., Dagar, S.S., et al. (2014). Anaerobic fungi (phylum Neocallimastigomycota): advances in

understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *Fems Microbiology Ecology* 90(1), 1-17. doi: 10.1111/1574-6941.12383.
Guan, L.L., Nkrumah, J.D., Basarab, J.A., and Moore, S.S. (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *Fems Microbiology Letters* 288(1), 85-91. doi: 10.1111/j.1574-6968.2008.01343.x.
Gunsett, F.C. (1984). Linear Index Selection to Improve Traits Defined as Ratios. *Journal of Animal Science* 59(5), 1185-1193. doi: 10.2527/jas1984.5951185x.
Gupta, A., Kumar, S., Prasoodanan, V.P.K., Harish, K., Sharma, A.K., and Sharma, V.K. (2016a). Reconstruction of Bacterial and Viral Genomes from Multiple Metagenomes. *Frontiers in Microbiology* 7, 469. doi: 10.3389/fmicb.2016.00469.
Gupta, A., Kumar, S., Prasoodanan, V.P.K., Harish, K., Sharma, A.K., and Sharma, V.K. (2016b). Reconstruction of Bacterial and Viral Genomes from Multiple Metagenomes. *Frontiers in Microbiology* 7(469). doi: 10.3389/fmicb.2016.00469.
Hackmann, T.J., and Firkins, J.L. (2015). Maximizing efficiency of rumen microbial protein production. *Frontiers in Microbiology* 6. doi: 10.3389/fmicb.2015.00465.
Hall, M.B. (2009). Determination of Starch, Including Maltooligosaccharides, in Animal Feeds: Comparison of Methods and a Method Recommended for AOAC Collaborative Study. *Journal of Aoac International* 92(1), 42-49.
Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology* 5(10), R245-R249. doi: https://doi.org/10.1016/S1074-5521(98)90108-9.
Hart, E.H., Creevey, C.J., Hitch, T., and Kingston-Smith, A.H. (2018). Meta-proteomics of rumen microbiota indicates niche compartmentalisation and functional dominance in a limited number of metabolic pathways between abundant bacteria. *Scientific Reports* 8(1), 10504. doi: 10.1038/s41598-018-28827-7.
Healey, G.R., Murphy, R., Brough, L., Butts, C.A., and Coad, J. (2017). Interindividual variability in gut microbiota and host response to dietary interventions. *Nutrition Reviews* 75(12), 1059-1080.
Hegarty, R.S., Goopy, J.P., Herd, R.M., and McCorkell, B. (2007). Cattle selected for lower residual feed intake have reduced daily methane production. *J Anim Sci* 85(6), 1479-1486. doi: 10.2527/jas.2006-236.
Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., and Janssen, P.H. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci Rep* 5, 14567. doi: 10.1038/srep14567.
Hernandez-Sanabria, E., Goonewardene, L.A., Wang, Z., Durunna, O.N., Moore, S.S., and Guan, L.L. (2012). Impact of Feed Efficiency and Diet on Adaptive Variations in the Bacterial Community in the Rumen Fluid of Cattle. *Applied and Environmental Microbiology* 78(4), 1203-1214. doi: 10.1128/aem.05114-11.
Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., et al. (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* 331(6016), 463-467. doi: 10.1126/science.1200387.

Himmel, M.E., Ding, S.-Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W., et al. (2007). Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science* (5813).

Ho, Y.W., Abdullah, N., and Jalaludin, S. (1988). Penetrating structures of anaerobic rumen fungi in cattle and swamp buffalo. *Journal of General Microbiology* 134(1)**,** 177-182.

Ho, Y.W., and Barr, D.J.S. (1995). Classification of anaerobic gut fungi from herbivores with emphasis on rumen fungi from malaysia. *Mycologia* 87(5)**,** 655-677. doi: 10.2307/3760810.

Hobson, P.N., and Stewart, C.S. (1997). *The Rumen Microbial Ecosystem.* London: Blackie Academic & Professional.

Hogg, D., Pell, G., Dupree, P., Goubet, F., Martín-Orúe, S.M., Armand, S., et al. (2003). The modular architecture of Cellvibrio japonicus mannanases in glycoside hydrolase families 5 and 26 points to differences in their role in mannan degradation. *The Biochemical journal* 371(Pt 3)**,** 1027-1043. doi: 10.1042/BJ20021860.

Hristov, A.N., Callaway, T.R., Lee, C., and Dowd, S.E. (2012). Rumen bacterial, archaeal, and fungal diversity of dairy cows in response to ingestion of lauric or myristic acid. *J Anim Sci* 90(12)**,** 4449-4457. doi: 10.2527/jas.2011-4624.

Hristov, A.N., Oh, J., Firkins, J.L., Dijkstra, J., Kebreab, E., Waghorn, G., et al. (2013). SPECIAL TOPICS — Mitigation of methane and nitrous oxide emissions from animal operations: I. A review of enteric methane mitigation options. *Journal of Animal Science* 91**,** 5045-5069. doi: 10.2527/jas.2013-6583.

Hristov, A.N., Oh, J., Giallongo, F., Frederick, T.W., Harper, M.T., Weeks, H.L., et al. (2015). An inhibitor persistently decreased enteric methane emission from dairy cows with no negative effect on milk production. *Proceedings of the National Academy of Sciences* 112(34)**,** 10663-10668. doi: 10.1073/pnas.1504124112.

Hungate, R.E. (1966). *The rumen and its microbes.* New York and London: Academic Press.

Hungate, R.E. (1967). Hydrogen as an intermediate in the rumen fermentation. *Archiv für Mikrobiologie* 59(1-3)**,** 158-164. doi: 10.1007/BF00406327.

Huws, S.A., Creevey, C.J., Oyama, L.B., Mizrahi, I., Denman, S.E., Popova, M., et al. (2018). Addressing Global Ruminant Agricultural Challenges Through Understanding the Rumen Microbiome: Past, Present, and Future. *Frontiers in Microbiology* 9(2161). doi: 10.3389/fmicb.2018.02161.

Imelfort, M., Parks, D., Woodcroft, B.J., Dennis, P., Hugenholtz, P., and Tyson, G.W. (2014). GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ* 2**,** e603. doi: 10.7717/peerj.603.

IPCC (2006). *Guidelines for National Greenhouse Gas Inventories.* Japan: Institute for Global Environmental Strategies (IGES).

Jami, E., Israel, A., Kotser, A., and Mizrahi, I. (2013). Exploring the bovine rumen bacterial community from birth to adulthood. *Isme j* 7(6)**,** 1069-1079. doi: 10.1038/ismej.2013.2.

Jami, E., and Mizrahi, I. (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS One* 7(3)**,** e33306. doi: 10.1371/journal.pone.0033306.

Janssen, P.H. (2010). Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. *Animal Feed Science and Technology* 160(1-2), 1-22. doi: 10.1016/j.anifeedsci.2010.07.002.

Janssen, P.H., and Kirs, M. (2008a). Structure of the archaeal community of the rumen. *Appl Environ Microbiol* 74(12), 3619-3625. doi: 10.1128/aem.02812-07.

Janssen, P.H., and Kirs, M. (2008b). Structure of the Archaeal Community of the Rumen. *Applied & Environmental Microbiology* 74(12), 3619-3625. doi: 10.1128/AEM.02812-07.

Jayanegara, A., Sarwono, K.A., Kondo, M., Matsui, H., Ridla, M., Laconi, E.B., et al. (2018). Use of 3-nitrooxypropanol as feed additive for mitigating enteric methane emissions from ruminants: a meta-analysis AU - Jayanegara, Anuraga. *Italian Journal of Animal Science* 17(3), 650-656. doi: 10.1080/1828051X.2017.1404945.

Jin, D., Zhao, S., Zheng, N., Bu, D., Beckers, Y., Denman, S.E., et al. (2017). Differences in Ureolytic Bacterial Composition between the Rumen Digesta and Rumen Wall Based on ureC Gene Classification. *Frontiers in Microbiology* 8(385). doi: 10.3389/fmicb.2017.00385.

Johnson, D.E., Ferrell, C.L., and Jenkins, T.G. (2003). The history of energetic efficiency research: Where have we been and where are we going? *Journal of Animal Science* 81(13_suppl_1), E27-E38. doi: 10.2527/2003.8113_suppl_1E27x.

Johnson, K.A., and Johnson, D.E. (1995). Methane emissions from cattle. *Journal of Animal Science* 73(8), 2483-2492.

Jouany, J.P. (1995). Rumen microbial metabolism and ruminant digestion: J-P. Jouany (Ed). 1991. INRA, 147 rue de l'Université, 75338 Paris cedex 07, France. ISBN 2 7380 0345 1. 374 pp. Text in English. 350 FF. *Livestock Production Science* 44(3), 284. doi: https://doi.org/10.1016/0301-6226(95)90065-9.

Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., et al. (2016). Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4(1), 56-56.

Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3, e1165. doi: 10.7717/peerj.1165.

Kenters, N., Henderson, G., Jeyanathan, J., Kittelmann, S., and Janssen, P.H. (2011). Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. *Journal of Microbiological Methods* 84(1), 52-60. doi: https://doi.org/10.1016/j.mimet.2010.10.011.

Kim, M., Yu, Z., and Morrison, M. (2011). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiology Ecology* 76(1), 49-63. doi: 10.1111/j.1574-6941.2010.01029.x.

Kittelmann, S., Devente, S.R., Kirk, M.R., Seedorf, H., Dehority, B.A., and Janssen, P.H. (2015a). Phylogeny of Intestinal Ciliates, Including Charonina ventriculi, and Comparison of Microscopy and 18S rRNA Gene Pyrosequencing for Rumen Ciliate Community Structure Analysis. *Applied and Environmental Microbiology* 81(7), 2433. doi: 10.1128/AEM.03697-14.

Kittelmann, S., Devente, S.R., Kirk, M.R., Seedorf, H., Dehority, B.A., and Janssen, P.H. (2015b). Phylogeny of Intestinal Ciliates, Including Charonina ventriculi, and Comparison of Microscopy and 18S rRNA Gene Pyrosequencing for Rumen Ciliate Community Structure Analysis. *Appl Environ Microbiol* 81(7), 2433-2444. doi: 10.1128/aem.03697-14.

Kittelmann, S., Pinares-Patiño, C.S., Seedorf, H., Kirk, M.R., Ganesh, S., McEwan, J.C., et al. (2014). Two Different Bacterial Community Types Are Linked with the Low-Methane Emission Trait in Sheep. *PLoS ONE* 9(7), 1-9. doi: 10.1371/journal.pone.0103171.

Kittelmann, S., Seedorf, H., Walters, W.A., Clemente, J.C., Knight, R., Gordon, J.I., et al. (2013a). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS One* 8(2), e47879. doi: 10.1371/journal.pone.0047879.

Kittelmann, S., Seedorf, H., Walters, W.A., Clemente, J.C., Knight, R., Gordon, J.I., et al. (2013b). Simultaneous Amplicon Sequencing to Explore Co-Occurrence Patterns of Bacterial, Archaeal and Eukaryotic Microorganisms in Rumen Microbial Communities. *Plos One* 8(2). doi: 10.1371/journal.pone.0047879.

Knapp, J.R., Laur, G.L., Vadas, P.A., Weiss, W.P., and Tricarico, J.M. (2014). Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. *Journal of Dairy Science* 97(6), 3231-3261. doi: http://dx.doi.org/10.3168/jds.2013-7234.

Koch, R.M., Swiger, L.A., Chambers, D., and Gregory, K.E. (1963). Efficiency of Feed Use in Beef Cattle. *Journal of Animal Science* 22(2), 486-494. doi: 10.2527/jas1963.222486x.

Koetschan, C., Kittelmann, S., Lu, J., Al-Halbouni, D., Jarvis, G.N., Mueller, T., et al. (2014). Internal Transcribed Spacer 1 Secondary Structure Analysis Reveals a Common Core throughout the Anaerobic Fungi (Neocallimastigomycota). *Plos One* 9(3). doi: 10.1371/journal.pone.0091928.

Kong, R.S., Liang, G., Chen, Y., Stothard, P., and Guan le, L. (2016). Transcriptome profiling of the rumen epithelium of beef cattle differing in residual feed intake. *BMC Genomics* 17, 592. doi: 10.1186/s12864-016-2935-4.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics (Oxford, England)* 28(24), 3211-3217. doi: 10.1093/bioinformatics/bts611.

Korver, S. (1988). Genetic aspects of feed intake and feed efficiency in dairy cattle: A review. *Livestock Production Science* 20(1), 1-13. doi: https://doi.org/10.1016/0301-6226(88)90049-8.

Kosugi, A., Murashima, K., and Doi, R.H. (2001). Characterization of Xylanolytic Enzymes in Clostridium cellulovorans: Expression of Xylanase Activity Dependent on Growth Substrates. *Journal of Bacteriology* 183(24), 7037-7043. doi: 10.1128/jb.183.24.7037-7043.2001.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied And Environmental Microbiology* 79(17), 5112-5120. doi: 10.1128/AEM.01043-13.

Krause, D.O., Nagaraja, T.G., Wright, A.D.G., and Callaway, T.R. (2013). Board-invited review: Rumen microbiology: Leading the way in microbial ecology1,2. *Journal of Animal Science* 91(1), 331-341. doi: 10.2527/jas.2012-5567.

Kumar, S., Choudhury, P.K., Carro, M.D., Griffith, G.W., Dagar, S.S., Puniya, M., et al. (2014). New aspects and strategies for methane mitigation from ruminants. *Appl Microbiol Biotechnol* 98(1), 31-44. doi: 10.1007/s00253-013-5365-0.

Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D., and Charles, T.C. (2015). Current and future resources for functional metagenomics. *Frontiers in Microbiology* 6(1196). doi: 10.3389/fmicb.2015.01196.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology* 31, 814. doi: 10.1038/nbt.2676 https://www.nature.com/articles/nbt.2676 - supplementary-information.

Larance, M., and Lamond, A.I. (2015). Multidimensional proteomics for cell biology. *Nature Reviews Molecular Cell Biology* 16, 269. doi: 10.1038/nrm3970 https://www.nature.com/articles/nrm3970 - supplementary-information.

Leahy, S.C., Kelly, W.J., Altermann, E., Ronimus, R.S., Yeoman, C.J., Pacheco, D.M., et al. (2010). The Genome Sequence of the Rumen Methanogen Methanobrevibacter ruminantium Reveals New Possibilities for Controlling Ruminant Methane Emissions. *Plos One* 5(1). doi: 10.1371/journal.pone.0008926.

Lee, S.S., Ha, J.K., and Cheng, K.J. (2000). Relative contributions of bacteria, protozoa, and fungi to in vitro degradation of orchard grass cell walls and their interactions. *Applied and Environmental Microbiology* 66(9), 3807-3813. doi: 10.1128/aem.66.9.3807-3813.2000.

Li, F., and Guan, L.L. (2017a). Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Applied And Environmental Microbiology*. doi: 10.1128/AEM.00061-17.

Li, F., and Guan, L.L. (2017b). Metatranscriptomic Profiling Reveals Linkages between the Active Rumen Microbiome and Feed Efficiency in Beef Cattle. *Appl Environ Microbiol* 83(9). doi: 10.1128/aem.00061-17.

Li, F., Henderson, G., Sun, X., Cox, F., Janssen, P.H., and Guan le, L. (2016a). Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches. *Front Microbiol* 7, 987. doi: 10.3389/fmicb.2016.00987.

Li, F., Henderson, G., Sun, X., Cox, F., Janssen, P.H., and Guan, L.L. (2016b). Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches. *Frontiers in Microbiology* 7(987). doi: 10.3389/fmicb.2016.00987.

Li, F., Hitch, T.C.A., Chen, Y., Creevey, C.J., and Guan, L.L. (2019). Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the rumen microbiome and its associations with feed efficiency in beef cattle. *Microbiome* 7(1), 6. doi: 10.1186/s40168-019-0618-5.

Li, M., Penner, G.B., Hernandez-Sanabria, E., Oba, M., and Guan, L.L. (2009). Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen. *Journal of Applied Microbiology* 107(6), 1924-1934. doi: 10.1111/j.1365-2672.2009.04376.x.

Li, R.W., Connor, E.E., Li, C., Baldwin Vi, R.L., and Sparks, M.E. (2012). Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. *Environ Microbiol* 14(1), 129-139. doi: 10.1111/j.1462-2920.2011.02543.x.

Li, Z.P., Wright, A.D.G., Si, H.Z., Wang, X.X., Qian, W.X., Zhang, Z.G., et al. (2016c). Changes in the rumen microbiome and metabolites reveal the effect of host genetics on hybrid crosses. *Environmental Microbiology Reports* 8(6), 1016-1023. doi: 10.1111/1758-2229.12482.

Lin, C., Stahl, D.A., and Raskin, L. (1997). Microbial community structure in gastrointestinal tracts of domestic animals: comparative analyses using rRNA-targeted oligonucleotide probes. *FEMS Microbiology Ecology* 22(4), 281-294. doi: 10.1111/j.1574-6941.1997.tb00380.x.

Liquet, B., Cao, K.-A.L., Hocini, H., and Thiébaut, R. (2012). A novel approach for biomarker selection and the integration of repeated measures experiments from two assays. *BMC Bioinformatics* 13(1), 325. doi: 10.1186/1471-2105-13-325.

Liu, J.h., Zhang, M.l., Zhang, R.y., Zhu, W.y., and Mao, S.y. (2016). Comparative studies of the composition of bacterial microbiota associated with the ruminal content, ruminal epithelium and in the faeces of lactating dairy cows. *Microbial Biotechnology* 9(2), 257-268. doi: doi:10.1111/1751-7915.12345.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42(D1), D490-D495. doi: 10.1093/nar/gkt1178.

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

Lovell, D., Pawlowsky-Glahn, V., Egozcue, J.J., Marguerat, S., and Bähler, J. (2015). Proportionality: A Valid Alternative to Correlation for Relative Data. *PLoS Computational Biology* 11(3), 1-12. doi: 10.1371/journal.pcbi.1004075.

Malmuthuge, N., Li, M., Chen, Y., Fries, P., Griebel, P.J., Baurhoo, B., et al. (2012). Distinct commensal bacteria associated with ingesta and mucosal epithelium in the gastrointestinal tracts of calves and chickens. *Fems Microbiology Ecology* 79(2), 337-347. doi: 10.1111/j.1574-6941.2011.01220.x.

Mandal, S., Van Treuren, W., White, R.A., Eggesbo, M., Knight, R., and Peddada, S.D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease* 26, 27663-Article No.: 27663. doi: 10.3402/mehd.v26.27663.

Marchesi, J.R., and Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome* 3(1), 31. doi: 10.1186/s40168-015-0094-5.

Martin, C., Morgavi, D.P., and Doreau, M. (2010). Methane mitigation in ruminants: from microbe to the farm scale. *Animal* 4(3), 351-365. doi: 10.1017/s1751731109990620.

Martínez, I., Jaehyoung, K., Duffy, P.R., Schlegel, V.L., and Walter, J. (2010). Resistant Starches Types 2 and 4 Have Differential Effects on the Composition of the Fecal Microbiota in Human Subjects. *PLoS ONE* 5(11), 1-11. doi: 10.1371/journal.pone.0015046.

Martinez-Fernandez, G., Denman, S.E., Cheung, J., and McSweeney, C.S. (2017). Phloroglucinol Degradation in the Rumen Promotes the Capture of Excess

Hydrogen Generated from Methanogenesis Inhibition. *Frontiers in Microbiology* 8(1871). doi: 10.3389/fmicb.2017.01871.
Martinez-Fernandez, G., Denman, S.E., Yang, C., Cheung, J., Mitsumori, M., and McSweeney, C.S. (2016). Methane Inhibition Alters the Microbial Community, Hydrogen Flow, and Fermentation Response in the Rumen of Cattle. *Frontiers in Microbiology* 7(1122). doi: 10.3389/fmicb.2016.01122.
Matthews, C., Crispie, F., Lewis, E., Reid, M., O'Toole, P.W., and Cotter, P.D. (2018). The rumen microbiome: a crucial consideration when optimising milk and meat production and nitrogen utilisation efficiency AU - Matthews, Chloe. *Gut Microbes*, 1-18. doi: 10.1080/19490976.2018.1505176.
McAllister, T.A., Bae, H.D., Jones, G.A., and Cheng, K.J. (1994a). Microbial attachment and feed digestion in the rumen. *Journal of Animal Science* 72(11), 3004-3018.
McAllister, T.A., Bae, H.D., Yanke, L.J., Cheng, K.J., and Muir, A. (1994b). Effect of condensed tannins from birdsfoot trefoil on endoglucanase activity and the digestion of cellulose filter paper by ruminal fungi. *Can J Microbiol* 40(4), 298-305.
McAllister, T.A., Meale, S.J., Valle, E., Guan, L.L., Zhou, M., Kelly, W.J., et al. (2015). RUMINANT NUTRITION SYMPOSIUM: Use of genomics and transcriptomics to identify strategies to lower ruminal methanogenesis. *J Anim Sci* 93(4), 1431-1449. doi: 10.2527/jas.2014-8329.
McCabe, M.S., Cormican, P., Keogh, K., O'Connor, A., O'Hara, E., Palladino, R.A., et al. (2015a). Illumina MiSeq Phylogenetic Amplicon Sequencing Shows a Large Reduction of an Uncharacterised Succinivibrionaceae and an Increase of the Methanobrevibacter gottschalkii Clade in Feed Restricted Cattle. *PLoS One* 10(7), e0133234. doi: 10.1371/journal.pone.0133234.
McCabe, M.S., Cormican, P., Keogh, K., O'Connor, A., O'Hara, E., Palladino, R.A., et al. (2015b). Illumina MiSeq Phylogenetic Amplicon Sequencing Shows a Large Reduction of an Uncharacterised Succinivibrionaceae and an Increase of the Methanobrevibacter gottschalkii Clade in Feed Restricted Cattle. *PLOS ONE* 10(7), e0133234. doi: 10.1371/journal.pone.0133234.
McCann, J.C., Elolimy, A.A., and Loor, J.J. (2017). Rumen Microbiome, Probiotics, and Fermentation Additives. *The Veterinary Clinics Of North America. Food Animal Practice*. doi: 10.1016/j.cvfa.2017.06.009.
McGovern, E., McCabe, M.S., Cormican, P., Popova, M., Keogh, K., Kelly, A.K., et al. (2017). Plane of nutrition affects the phylogenetic diversity and relative abundance of transcriptionally active methanogens in the bovine rumen. *Scientific Reports* 7(1), 13047. doi: 10.1038/s41598-017-13013-y.
McMurdie, P.J., and 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.
McNeil, M., Darvill, A.G., Fry, S.C., and Albersheim, P. (1984). Structure and Function of the Primary Cell Walls of Plants. *Annual Review of Biochemistry* 53(1), 625-663. doi: 10.1146/annurev.bi.53.070184.003205.
McSweeney, C., and Mackie, R.I. (2012). *Microorganisms and ruminant digestion: state of knowledge, trends and future prospects.*
Mesuere, B., Devreese, B., Debyser, G., Aerts, M., Vandamme, P., and Dawyndt, P. (2012). Unipept: Tryptic Peptide-Based Biodiversity Analysis of Metaproteome

Samples. *Journal of Proteome Research* 11(12), 5773-5780. doi: 10.1021/pr300576s.
Miron, J., Ben-Ghedalia, D., and Morrison, M. (2001). Invited Review: Adhesion Mechanisms of Rumen Cellulolytic Bacteria. *Journal of Dairy Science* 84(6), 1294-1309. doi: 10.3168/jds.S0022-0302(01)70159-2.
Montanholi, Y.R., Swanson, K.C., Palme, R., Schenkel, F.S., McBride, B.W., Lu, D., et al. (2010). Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. *Animal* 4(5), 692-701. doi: 10.1017/s1751731109991522.
Morgavi, D., Kelly, W.J., Janssen, P.H., and Attwood, G.T. (2013). "Rumen microbial (meta) genomics and its application to ruminant production".).
Morgavi, D.P., Forano, E., Martin, C., and Newbold, C.J. (2010). Microbial ecosystem and methanogenesis in ruminants. *Animal* 4(7), 1024-1036. doi: 10.1017/s1751731110000546.
Morgavi, D.P., Kelly, W.J., Janssen, P.H., and Attwood, G.T. (2012a). Rumen microbial (meta)genomics and its application to ruminant production. *animal* 7(s1), 184-201. doi: 10.1017/S1751731112000419.
Morgavi, D.P., Martin, C., Jouany, J.P., and Ranilla, M.J. (2012b). Rumen protozoa and methanogenesis: not a simple cause-effect relationship. *Br J Nutr* 107(3), 388-397. doi: 10.1017/s0007114511002935.
Morrison, M., Murray, R.M., and Boniface, A.N. (1990). Nutrient metabolism and rumen microorganisms in sheep fed a poor-quality tropical grass hay supplemented with sulfate. *Journal of Agricultural Science* 115, 269-275.
Myer, P.R., Smith, T.P.L., Wells, J.E., Kuehn, L.A., and Freetly, H.C. (2015a). Rumen Microbiome from Steers Differing in Feed Efficiency. *PLOS ONE* 10(6), e0129174. doi: 10.1371/journal.pone.0129174.
Myer, P.R., Smith, T.P.L., Wells, J.E., Kuehn, L.A., and Freetly, H.C. (2015b). Rumen Microbiome from Steers Differing in Feed Efficiency. *Plos One* 10(6). doi: 10.1371/journal.pone.0129174.
Nations, U. (2015). *World Population Prospects* [Online]. [Accessed].
Network, I.F.C. (2016). "Long-term dairy outlook", in: *IFCN - the dairy research network*.).
Neves, A.L.A., Li, F., Ghoshal, B., McAllister, T., and Guan, L.L. (2017). Enhancing the Resolution of Rumen Microbial Classification from Metatranscriptomic Data Using Kraken and Mothur. *Frontiers in Microbiology* 8(2445). doi: 10.3389/fmicb.2017.02445.
Newbold, C.J., de la Fuente, G., Belanche, A., Ramos-Morales, E., and McEwan, N.R. (2015). The Role of Ciliate Protozoa in the Rumen. *Front Microbiol* 6, 1313. doi: 10.3389/fmicb.2015.01313.
Nocker, A., Burr, M., and Camper, A.K. (2007). Genotypic Microbial Community Profiling: A Critical Technical Review. *Microbial Ecology* 54(2), 276-289. doi: 10.1007/s00248-006-9199-5.
Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., et al. (2014). The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic acids research* 42(Database issue), D26-D31. doi: 10.1093/nar/gkt1069.

NRC (2016). *Nutrient Requirements of Beef Cattle.* Washington, D.C., USA: National Academy Press.
Olfert, E., Cross, B., and McWilliams, A. (1993). *Guide to the Care and Use of Experimental Steers.* Ottawa, ON, Canada: Canadian Council on Animal Care.
Orpin, C.G. (1975). Studies on rumen flagellate neocallimastix-frontalis. *Journal of General Microbiology* 91(DEC)**,** 249-262.
Orpin, C.G., and Joblin, K.N. (1997). "The rumen anaerobic fungi," in *The Rumen Microbial Ecosystem,* eds. P.N. Hobson & C.S. Stewart. Springer Netherlands), 140-195.
Ozkose, E., Thomas, B.J., Davies, D.R., Griffith, G.W., and Theodorou, M.K. (2001). Cyllamyces aberensis gen.nov sp.nov., a new anaerobic gut fungus with branched sporangiophores isolated from cattle. *Canadian Journal of Botany-Revue Canadienne De Botanique* 79(6)**,** 666-673.
Palackal, N., Lyon, C.S., Zaidi, S., Luginbühl, P., Dupree, P., Goubet, F., et al. (2007). A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Applied Microbiology and Biotechnology* 74(1)**,** 113-124. doi: 10.1007/s00253-006-0645-6.
Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25(7)**,** 1043-1055. doi: 10.1101/gr.186072.114.
Parmar, N.R., Solanki, J.V., Patel, A.B., Shah, T.M., Patel, A.K., Parnerkar, S., et al. (2014). Metagenome of Mehsani buffalo rumen microbiota: an assessment of variation in feed-dependent phylogenetic and functional classification. *J Mol Microbiol Biotechnol* 24(4)**,** 249-261. doi: 10.1159/000365054.
Patil, K.R., Haider, P., Pope, P.B., Turnbaugh, P.J., Morrison, M., Scheffer, T., et al. (2011). Taxonomic metagenome sequence assignment with structured output models. *Nature Methods* 8**,** 191. doi: 10.1038/nmeth0311-191
https://www.nature.com/articles/nmeth0311-191 - supplementary-information.
Patra, A., Park, T., Kim, M., and Yu, Z. (2017). Rumen methanogens and mitigation of methane emission by anti-methanogenic compounds and substances. *Journal of Animal Science and Biotechnology* 8(1)**,** 13. doi: 10.1186/s40104-017-0145-9.
Paul, S.S., Yu, Z., Bu, D., Xu, J., and Hyde, K.D. (2018). A phylogenetic census of global diversity of gut anaerobic fungi and a new taxonomic framework. *Fungal Diversity* 89(1)**,** 253-266. doi: 10.1007/s13225-018-0396-6.
Pawlowsky- Glahn, V., Egozcue, J.J., and Tolosana- Delgado, R. (2015). *Modelling and Analysis of Compositional Data.*
Pearson, K. (1897). Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London* 60(359-367)**,** 489-498. doi: 10.1098/rspl.1896.0076.
Petri, R.M., Schwaiger, T., Penner, G.B., Beauchemin, K.A., Forster, R.J., McKinnon, J.J., et al. (2013). Characterization of the Core Rumen Microbiome in Cattle during Transition from Forage to Concentrate as Well as during and after an Acidotic Challenge. *Plos One* 8(12). doi: 10.1371/journal.pone.0083424.

Pickering, N.K., Oddy, V.H., Basarab, J., Cammack, K., Hayes, B., Hegarty, R.S., et al. (2015a). Animal board invited review: genetic possibilities to reduce enteric methane emissions from ruminants. *Animal* 9(9), 1431-1440. doi: 10.1017/s1751731115000968.

Pickering, N.K., Oddy, V.H., Basarab, J., Cammack, K., Hayes, B., Hegarty, R.S., et al. (2015b). Animal board invited review: genetic possibilities to reduce enteric methane emissions from ruminants. *Animal : an international journal of animal bioscience* 9(9), 1431-1440. doi: 10.1017/S1751731115000968.

Poelaert, C., Nollevaux, G., Boudry, C., Taminiau, B., Nezer, C., Daube, G., et al. (2017). Reducing agent can be omitted in the incubation medium of the batch in vitro fermentation model of the pig intestines. *animal* 12(6), 1154-1164. doi: 10.1017/S1751731117002749.

Poretsky, R., Rodriguez, R.L., Luo, C., Tsementzi, D., and Konstantinidis, K.T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One* 9(4), e93827. doi: 10.1371/journal.pone.0093827.

Poulsen, M., Schwab, C., Borg Jensen, B., Engberg, R.M., Spang, A., Canibe, N., et al. (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nature Communications* 4, 1428. doi: 10.1038/ncomms2432

https://www.nature.com/articles/ncomms2432 - supplementary-information.

Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., et al. (2011a). Snapshot of the Eukaryotic Gene Expression in Muskoxen Rumen-A Metatranscriptomic Approach. *Plos One* 6(5). doi: 10.1371/journal.pone.0020521.

Qi, M., Wang, P., Selinger, L.B., Yanke, L.J., Forster, R.J., and McAllister, T.A. (2011b). Isolation and characterization of a ferulic acid esterase (Fae1A) from the rumen fungus Anaeromyces mucronatus. *Journal of Applied Microbiology* 110(5), 1341-1350. doi: 10.1111/j.1365-2672.2011.04990.x.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41(Database issue), D590-D596. doi: 10.1093/nar/gks1219.

Ramírez, L., Arrizon, J., Sandoval, G., Cardador, A., Bello-Mendoza, R., Lappe, P., et al. (2008). A New Microplate Screening Method for the Simultaneous Activity Quantification of Feruloyl Esterases, Tannases, and Chlorogenate Esterases. *Applied Biochemistry and Biotechnology* 151(2), 711-723. doi: 10.1007/s12010-008-8319-8.

Rashamuse, K.J., Burton, S.G., and Cowan, D.A. (2007). A novel recombinant ethyl ferulate esterase from Burkholderia multivorans. *Journal of Applied Microbiology* 103(5), 1610-1620. doi: 10.1111/j.1365-2672.2007.03394.x.

Ribeiro, G.O., Gruninger, R.J., Badhan, A., and McAllister, T.A. (2016). Mining the rumen for fibrolytic feed enzymes. *Animal Frontiers* 6(2), 20-26. doi: 10.2527/af.2016-0019.

Rico, D.E., Preston, S.H., Risser, J.M., and Harvatine, K.J. (2015). Rapid changes in key ruminal microbial populations during the induction of and recovery from diet-

induced milk fat depression in dairy cows. *British Journal of Nutrition* 114(3), 358-367. doi: 10.1017/s0007114515001865.
Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1), 139-140. doi: 10.1093/bioinformatics/btp616.
Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11(3), R25. doi: 10.1186/gb-2010-11-3-r25.
Rodríguez, H., Curiel, J.A., Landete, J.M., de las Rivas, B., de Felipe, F.L., Gómez-Cordovés, C., et al. (2009). Food phenolics and lactic acid bacteria. *International Journal of Food Microbiology* 132(2), 79-90. doi: https://doi.org/10.1016/j.ijfoodmicro.2009.03.025.
Roehe, R., Dewhurst, R.J., Duthie, C.-A., Rooke, J.A., McKain, N., Ross, D.W., et al. (2016). Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. *Plos Genetics* 12(2). doi: 10.1371/journal.pgen.1005846.
Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology* 13(11), 1-19. doi: 10.1371/journal.pcbi.1005752.
Rojas-Downing, M.M., Nejadhashemi, A.P., Harrigan, T., and Woznicki, S.A. (2017). Climate change and livestock: Impacts, adaptation, and mitigation. *Climate Risk Management* 16, 145-163. doi: https://doi.org/10.1016/j.crm.2017.02.001.
Romain, G., Thomas, T., and Margaret, E.K.E. (2009). Ecological Conditions Affect Evolutionary Trajectory in a Predator-Prey System. *Evolution* (3), 641. doi: 10.1111/j.1558-5646.2008.00559.x.
Romero-Perez, A., Okine, E.K., McGinn, S.M., Guan, L.L., Oba, M., Duval, S.M., et al. (2014). The potential of 3-nitrooxypropanol to lower enteric methane emissions from beef cattle. *Journal of Animal Science* 92(10), 4682-4693. doi: 10.2527/jas.2014-7573.
Rosnow, J.J., Anderson, L.N., Nair, R.N., Baker, E.S., and Wright, A.T. (2017). Profiling microbial lignocellulose degradation and utilization by emergent omics technologies. *Critical Reviews in Biotechnology* 37(5), 626-640. doi: 10.1080/07388551.2016.1209158.
Russell, J.B., and Rychlik, J.L. (2001). Factors that alter rumen microbial ecology. *Science* 292(5519), 1119-1122. doi: 10.1126/science.1058830.
Sabree, Z.L., Rondon, M.R., and Handelsman, J. (2009). "Metagenomics," in *Encyclopedia of Microbiology (Third Edition),* ed. M. Schaechter. (Oxford: Academic Press), 622-632.
Saha, B.C. (2000). α-l-Arabinofuranosidases: biochemistry, molecular biology and application in biotechnology. *Biotechnology Advances* 18(5), 403-423. doi: https://doi.org/10.1016/S0734-9750(00)00044-6.
Saha, B.C. (2003). Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology* 30(5), 279-291. doi: 10.1007/s10295-003-0049-x.
Salonen, A., Lahti, L., Salojärvi, J., Holtrop, G., Korpela, K., Duncan, S.H., et al. (2014). Impact of diet and individual variation on intestinal microbiota composition and

fermentation products in obese men. *ISME Journal: Multidisciplinary Journal of Microbial Ecology* 8(11), 2218-2230. doi: 10.1038/ismej.2014.63.
Sanchez, C. (2009). Lignocellulosic residues: Biodegradation and bioconversion by fungi. *Biotechnology Advances* 27(2), 185-194. doi: 10.1016/j.biotechadv.2008.11.001.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied & Environmental Microbiology* 75(23), 7537-7541. doi: 10.1128/AEM.01541-09.
Seedorf, H., Kittelmann, S., Henderson, G., and Janssen, P.H. (2014). RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. *Peerj* 2. doi: 10.7717/peerj.494.
Sehgal, J.P., Jit, D., Puniya, A.K., and Singh, K. (2008). Influence of anaerobic fungal administration on growth, rumen fermentation and nutrient digestion in female buffalo calves. *Journal of Animal and Feed Sciences* 17(4), 510-518. doi: 10.22358/jafs/66678/2008.
Seshadri, R., Leahy, S.C., Attwood, G.T., Teh, K.H., Lambie, S.C., Cookson, A.L., et al. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nature Biotechnology* 36(4), 359-367. doi: 10.1038/nbt.4110.
Shaani, Y., Zehavi, T., Eyal, S., Miron, J., and Mizrahi, I. (2018). Microbiome niche modification drives diurnal rumen community assembly, overpowering individual variability and diet effects. *The ISME Journal*. doi: 10.1038/s41396-018-0203-0.
Shabat, S.K.B., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., Berg Miller, M.E., et al. (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J*. doi: 10.1038/ismej.2016.62.
Shallom, D., and Shoham, Y. (2003). Microbial hemicellulases. *Current Opinion in Microbiology* 6(3), 219-228. doi: https://doi.org/10.1016/S1369-5274(03)00056-0.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11), 2498-2504. doi: 10.1101/gr.1239303.
Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics* 14, 618. doi: 10.1038/nrg3542.
Silley, P. (1985). A note on the pectinolytic enzymes of Lachnospira multiparus. *Journal of Applied Bacteriology* 58(2), 145-149. doi: 10.1111/j.1365-2672.1985.tb01441.x.
Silverman, J.D., Washburne, A.D., Mukherjee, S., and David, L.A. (2017). A phylogenetic transform enhances analysis of compositional microbiota data. *eLife* 6, e21887. doi: 10.7554/eLife.21887.
Sirohi, S.K., Pandey, N., Singh, B., and Puniya, A.K. (2010). Rumen methanogens: a review. *Indian Journal of Microbiology* 50(3), 253-262. doi: 10.1007/s12088-010-0061-6.
Skene, I.K., and Brooker, J.D. (1995). Characterization of Tannin Acylhydrolase Activity in the Ruminal Bacterium Selenomonas Ruminantium. *Anaerobe* 1(6), 321-327. doi: https://doi.org/10.1006/anae.1995.1034.

Snelling, T.J., and Wallace, R.J. (2017). The rumen microbial metaproteome as revealed by SDS-PAGE. *BMC Microbiology* 17(1), 9. doi: 10.1186/s12866-016-0917-y.

Sniffen, C.J., and Robinson, P.H. (1987). Microbial growth and flow as influenced by dietary manipulations. *Journal of Dairy Science* 70(2), 425-441.

Stackebrandt, E., and Goebel, B.M. (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology* 44(4), 846-849. doi: doi:10.1099/00207713-44-4-846.

Steele, M.A., Penner, G.B., Chaucheyras-Durand, F., and Guan, L.L. (2016). Development and physiology of the rumen and the lower gut: Targets for improving gut health. *J Dairy Sci*. doi: 10.3168/jds.2015-10351.

Stevenson, D.M., and Weimer, P.J. (2007). Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Applied Microbiology and Biotechnology* 75(1), 165-174. doi: 10.1007/s00253-006-0802-y.

Stewart, C.S., Fonty, G., and Gouet, P. (1988). The establishment of rumen microbial communities. *Animal Feed Science and Technology* 21(2), 69-97. doi: https://doi.org/10.1016/0377-8401(88)90093-4.

Stewart, R.D., Warr, A., Watson, M., Auffret, M.D., Dewhurst, R.J., Roehe, R., et al. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nature Communications* 9(1). doi: 10.1038/s41467-018-03317-6.

Sunagawa, S., Mende, D.R., Zeller, G., Izquierdo-Carrasco, F., Berger, S.A., Kultima, J.R., et al. (2013a). Metagenomic species profiling using universal phylogenetic marker genes. 10(12), 1196-1199. doi: 10.1038/nmeth.2693.

Sunagawa, S., Mende, D.R., Zeller, G., Izquierdo-Carrasco, F., Berger, S.A., Kultima, J.R., et al. (2013b). Metagenomic species profiling using universal phylogenetic marker genes. *Nature Methods* 10(12), 1196-1199. doi: 10.1038/nmeth.2693.

Svartström, O., Alneberg, J., Terrapon, N., Lombard, V., de Bruijn, I., Malmsten, J., et al. (2017). Ninety-nine de novo assembled genomes from the moose (Alces alces) rumen microbiome provide new insights into microbial plant biomass degradation. *The ISME Journal*. doi: 10.1038/ismej.2017.108.

Sylvester, J.T., Karnati, S.K., Yu, Z., Morrison, M., and Firkins, J.L. (2004). Development of an assay to quantify rumen ciliate protozoal biomass in cows using real-time PCR. *J Nutr* 134(12), 3378-3384.

Tajima, K., Aminov, R.I., Nagamine, T., Matsui, H., Nakamura, M., and Benno, Y. (2001). Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR. *Applied and Environmental Microbiology* 67(6), 2766-2774. doi: 10.1128/aem.67.6.2766-2774.2001.

Tap, J., Furet, J.-P., Bensaada, M., Philippe, C., Roth, H., Rabot, S., et al. (2015). Gut microbiota richness promotes its stability upon increased dietary fibre intake in healthy adults. *Environmental Microbiology* 17(12), 4954-4964. doi: 10.1111/1462-2920.13006.

Tapio, I., Shingfield, K.J., McKain, N., Bonin, A., Fischer, D., Bayat, A.R., et al. (2016). Oral Samples as Non-Invasive Proxies for Assessing the Composition of the Rumen Microbial Community. *PLoS One* 11(3), e0151220. doi: 10.1371/journal.pone.0151220.

Tapio, I., Snelling, T.J., Strozzi, F., and Wallace, R.J. (2017). The ruminal microbiome associated with methane emissions from ruminant livestock. *J Anim Sci Biotechnol* 8**,** 7. doi: 10.1186/s40104-017-0141-0.

Tewoldebrhan, T.A., Appuhamy, J.A.D.R.N., Lee, J.J., Niu, M., Seo, S., Jeong, S., et al. (2017). Exogenous β-mannanase improves feed conversion efficiency and reduces somatic cell count in dairy cattle. *Journal of Dairy Science* 100(1)**,** 244-252. doi: https://doi.org/10.3168/jds.2016-11017.

Theodorou, M.K., Mennim, G., Davies, D.R., Zhu, W.-Y., Trinci, A.P.J., and Brookman, J.L. (2007). Anaerobic fungi in the digestive tract of mammalian herbivores and their potential for exploitation. *Proceedings of the Nutrition Society* 55(3)**,** 913-926. doi: 10.1079/PNS19960088.

Thompson, S. (2015). *The effect of diet type on residual feed intake and the use of infrared thermography as a method to predict efficiency in beef bulls.* M.Sc., University of Manitoba.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., et al. (2009). A core gut microbiome in obese and lean twins. *Nature* 457(7228)**,** 480-484. doi: http://www.nature.com/nature/journal/v457/n7228/suppinfo/nature07540_S1.html.

Tymensen, L.D., Beauchemin, K.A., and McAllister, T.A. (2012). Structures of free-living and protozoa-associated methanogen communities in the bovine rumen differ according to comparative analysis of 16S rRNA and mcrA genes. *Microbiology* 158(7)**,** 1808-1817. doi: doi:10.1099/mic.0.057984-0.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428**,** 37. doi: 10.1038/nature02340
https://www.nature.com/articles/nature02340 - supplementary-information.

Van Soest, P.J., Robertson, J.B., and Lewis, B.A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *Journal of Dairy Science* 74(10)**,** 3583-3597.

Vanwonterghem, I., Evans, P.N., Parks, D.H., Jensen, P.D., Woodcroft, B.J., Hugenholtz, P., et al. (2016). Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nature Microbiology* 1**,** 16170. doi: 10.1038/nmicrobiol.2016.170
http://www.nature.com/articles/nmicrobiol2016170 - supplementary-information.

Varga, G.A., and Kolver, E.S. (1997). Microbial and animal limitations to fiber digestion and utilization. *The Journal Of Nutrition* 127(5 Suppl)**,** 819S-823S. doi: 10.1093/jn/127.5.819S.

Veira, D.M. (1986). The role of ciliate protozoa in nutrition of the ruminant. *J Anim Sci* 63(5)**,** 1547-1560.

Verberkmoes, N.C., Russell, A.L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., et al. (2008). Shotgun metaproteomics of the human distal gut microbiota. *The Isme Journal* 3**,** 179. doi: 10.1038/ismej.2008.108
https://www.nature.com/articles/ismej2008108 - supplementary-information.

Větrovský, T., and Baldrian, P. (2013). The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE* 8(2), e57923. doi: 10.1371/journal.pone.0057923.
Wallace, R.J., Rooke, J.A., Duthie, C.-A., Hyslop, J.J., Ross, D.W., McKain, N., et al. (2014). Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. *Scientific Reports* 4. doi: 10.1038/srep05892.
Wallace, R.J., Rooke, J.A., McKain, N., Duthie, C.-A., Hyslop, J.J., Ross, D.W., et al. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 16, 1-14. doi: 10.1186/s12864-015-2032-0.
Wang, L., Hatem, A., Catalyurek, U.V., Morrison, M., and Yu, Z. (2013). Metagenomic Insights into the Carbohydrate-Active Enzymes Carried by the Microorganisms Adhering to Solid Digesta in the Rumen of Cows. *PLoS ONE* 8(11), 1-11. doi: 10.1371/journal.pone.0078507.
Wang, S., Giller, K., Kreuzer, M., Ulbrich, S.E., Braun, U., and Schwarm, A. (2017). Contribution of Ruminal Fungi, Archaea, Protozoa, and Bacteria to the Methane Suppression Caused by Oilseed Supplemented Diets. *Frontiers in Microbiology* 8(1864). doi: 10.3389/fmicb.2017.01864.
Wang, Y., and McAllister, T.A. (2002). Rumen Microbes, Enzymes and Feed Digestion-A Review. *Asian-Australas J Anim Sci* 15(11), 1659-1676. doi: 10.5713/ajas.2002.1659.
Warton, D.I., Wright, S.T., and Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution* 3(1), 89-101. doi: 10.1111/j.2041-210X.2011.00127.x.
Weimer, P.J. (1992). Cellulose degradation by ruminal microorganisms. *Critical Reviews in Biotechnology* 12, 189-223.
Weimer, P.J. (2015). Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. *Frontiers in Microbiology* 6(296). doi: 10.3389/fmicb.2015.00296.
Weimer, P.J., Lopezguisa, J.M., and French, A.D. (1990). Effect of cellulose fine-structure on kinetics of its digestion by mixed ruminal microorganisms invitro. *Applied and Environmental Microbiology* 56(8), 2421-2429.
Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5. doi: 10.1186/s40168-017-0237-y.
Westerhuis, J.A., van Velzen, E.J.J., Hoefsloot, H.C.J., and Smilde, A.K. (2010). Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics* 6(1), 119-128. doi: 10.1007/s11306-009-0185-z.
Wilkinson, T.J., Huws, S.A., Edwards, J.E., Kingston-Smith, A.H., Siu-Ting, K., Hughes, M., et al. (2018). CowPI: A Rumen Microbiome Focussed Version of the PICRUSt Functional Inference Software. *Frontiers in Microbiology* 9(1095). doi: 10.3389/fmicb.2018.01095.
Williams, A.G., and Coleman, G.S. (1997). "The rumen protozoa," in *The Rumen Microbial Ecosystem,* eds. P.N. Hobson & C.S. Stewart. (Dordrecht: Springer Netherlands), 73-139.
Wilmes, P., and Bond, P.L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic

microorganisms. *Environmental Microbiology* 6(9), 911-920. doi: doi:10.1111/j.1462-2920.2004.00687.x.
Wolff, S.M., Ellison, M.J., Hao, Y., Cockrum, R.R., Austin, K.J., Baraboo, M., et al. (2017). Diet shifts provoke complex and variable changes in the metabolic networks of the ruminal microbiome. *Microbiome* 5(1), 60. doi: 10.1186/s40168-017-0274-6.
Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15(3), R46-R46. doi: 10.1186/gb-2014-15-3-r46.
Wu, L., Wang, H., Zhang, Z., Lin, R., Zhang, Z., and Lin, W. (2011). Comparative Metaproteomic Analysis on Consecutively Rehmannia glutinosa-Monocultured Rhizosphere Soil. *PLOS ONE* 6(5), e20611. doi: 10.1371/journal.pone.0020611.
Yang, T., Wei, Z., Friman, V.-P., Xu, Y., Shen, Q., Kowalchuk, G.A., et al. (2017). Resource availability modulates biodiversity-invasion relationships by altering competitive interactions. *Environmental Microbiology* 19(8), 2984-2991. doi: 10.1111/1462-2920.13708.
Yi, L., Shi, S., Wang, Y., Huang, W., Xia, Z.-a., Xing, Z., et al. (2016). Serum Metabolic Profiling Reveals Altered Metabolic Pathways in Patients with Post-traumatic Cognitive Impairments. *Scientific Reports* 6, 21320. doi: 10.1038/srep21320.
Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *BioTechniques* 36(5), 808-812. doi: 10.2144/04365ST04.
Zehavi, T., Probst, M., and Mizrahi, I. (2018). Insights Into Culturomics of the Rumen Microbiome. *Frontiers in Microbiology* 9(1999). doi: 10.3389/fmicb.2018.01999.
Zhang, Y.H.P., Himmel, M.E., and Mielenz, J.R. (2006). Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances* (5), 452. doi: 10.1016/j.biotechadv.2006.03.003.
Zhou, M., Chung, Y.H., Beauchemin, K.A., Holtshausen, L., Oba, M., McAllister, T.A., et al. (2011). Relationship between rumen methanogens and methane production in dairy cows fed diets supplemented with a feed enzyme additive. *Journal of Applied Microbiology* 111(5), 1148-1158. doi: 10.1111/j.1365-2672.2011.05126.x.
Zhou, M., Hernandez-Sanabria, E., and Guan, L.L. (2009). Assessment of the Microbial Ecology of Ruminal Methanogens in Cattle with Different Feed Efficiencies. *Applied and Environmental Microbiology* 75(20), 6524-6533. doi: 10.1128/aem.02815-08.
Zhou, M., Hernandez-Sanabria, E., and Guan, L.L. (2010). Characterization of Variation in Rumen Methanogenic Communities under Different Dietary and Host Feed Efficiency Conditions, as Determined by PCR-Denaturing Gradient Gel Electrophoresis Analysis. *Applied and Environmental Microbiology* 76(12), 3776-3786. doi: 10.1128/aem.00010-10.
Zhou, M., Hünerberg, M., Beauchemin, K.A., McAllister, T.A., Okine, E.K., and Guan, L.L. (2012). Individuality of ruminal methanogen/protozoa populations in beef cattle fed diets containing dried distillers' grain with solubles. *Acta Agriculturae Scandinavica: Section A, Animal Science* 62(4), 273-288. doi: 10.1080/09064702.2013.788206.
Zhou, M., Peng, Y.-J., Chen, Y., Klinger, C.M., Oba, M., Liu, J.-X., et al. (2018). Assessment of microbiome changes after rumen transfaunation: implications on

improving feed efficiency in beef cattle. *Microbiome* 6(1), 62-62. doi: 10.1186/s40168-018-0447-y.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. (2009). "Mixed Effects Models and Extensions in Ecology with R," in *Mixed Effects Models and Extensions in Ecology with R*.).

Chapter 2

Dynamics of microbial populations driven by interactions between diet and host shed light on individualized rumen microbiota

2.1 Introduction

The bovine rumen harbors a symbiotic community of anaerobic bacteria, archaea, fungi, and protozoa that plays an essential role in feed degradation and energy provision for the host. The fermentation of host-indigestible plant biomass by the rumen microbiota provides up to 70% of the energetic requirements of ruminant animals, mainly through the production of VFAs (NRC 2016, Russell and Rychlik 2001). While vital to efficient milk and meat production, rumen microbial fermentation is also a significant source of greenhouse gas emissions (Huws et al. 2018), with CH_4 exhibiting a global warming potential 25 times greater than that of CO_2 (IPCC 2006), as well as reducing host dietary energy availability by up to 12% (Johnson and Johnson 1995). Microbial composition in the rumen is primarily affected by the diet and exhibits a large degree of individual variation that may be a reflection of differences in the host nutrient utilization efficiency (Henderson et al. 2015). Because individuals may respond differently to dietary changes (Hernandez-Sanabria et al. 2012), a greater understanding of the drivers that affect the relationship between the rumen microbiota and host responsiveness to the diet could help to develop better dietary interventions to meet individual host nutritional requirements (Healey et al. 2017, Morgavi et al. 2012a).

The success of rumen microbes in digesting recalcitrant feedstuffs depends on the microbial interactions that occur in the ruminal ecosystem (Wolff et al. 2017, Russell and Rychlik 2001), which is in turn modulated by at least two possible factors: 1) the environment (e.g., nutrient availability for microbial growth) (Yang et al. 2017), and 2) the inter-individual variability of the host organism in terms of feed degradation (Weimer 2015). While nutrient availability is related to niche occupancy (Yang et al. 2017), inter-individual variability may reveal the effectiveness of the microbiota of one particular animal versus another in terms of efficiency of feed utilization (Weimer 2015). Individual animals are heterogeneous regarding the responsiveness to the diet, and thus they

exhibit a ‘personalized’ rumen microbial composition even when fed an identical diet and raised in the same environment (Brulc et al. 2009, Z. P. Li et al. 2016). Such host specificity does not appear to be restricted only to bacteria, as the inter-individual variation has also been observed in the methanogenic archaeal and protozoal communities (Zhou et al. 2012). Despite the wealth of information explaining how individual hosts control the rumen microbiota in relation to phenotypic traits (e.g., feed efficiency and CH_4 emissions) (Roehe et al. 2016), little is known regarding the precise mechanisms of host-microbe interactions in ruminants undergoing dietary changes. Moreover, interactions of host variation in feed efficiency with rumen microbial populations are not well understood, and further investigating their interactive effects on animal productivity may contribute to the identification of feed efficient animals and the reduction of methane emissions.

In this context, several studies have demonstrated that rumen bacteria and archaea (composition and abundance) play important roles in the feed conversion efficiency of cattle (Fuyong Li and Le Luo Guan 2017, Neves et al. 2017), suggesting that the identification of host-specific microbes should be considered as the regulating targets to improve host feed efficiency. Sequencing technologies were indispensable tools to quantify transcripts involved in feed efficiency in those studies, but we speculated that the application of qPCR in the current study could offer an opportunity to assess the dynamics of the rumen microbiota in a cost-effective manner. Here, we hypothesized that the stratification of dietary responses obtained from the magnitude of change in baseline rumen microbiota using qPCR can be used to identify individualized shifts of rumen microbes in cattle undergoing diet changes. The objectives of this research were (i) to investigate the dynamics of rumen microorganisms in cattle raised under different feeding regimens (forage vs. grain) and (ii) understand the relationship among the abundance of these microorganisms, host individuality and phenotypic traits.

2.2 Materials and methods

2.2.1 Animal study

The experimental procedures described in this study were approved by the Veterinary Services and the Animal Care Committee, University of Manitoba, Canada. Fifty-nine

purebred Angus bulls (mean age of 249 ± 22 days and average body weight of 313.9 ± 32 kg) were raised in confinement at the Glenlea Research Station, University of Manitoba, Canada in accordance with the guidelines of the Canadian Council on Animal Care (Olfert et al. 1993). The bulls were randomly assigned into four pens, and each pen was bedded with a mixture of barley/flaxseed straw and equipped with GrowSafe® (GrowSafe Systems Ltd., Airdrie, Alberta) feed bunks and a heated watering bowl. Bulls were fed forage or grain diets over two experimental periods (Period 1 and 2, each with an 80-day duration) in a crossover design (Table S1). In the first feeding period (FP1), two pens (1 and 2) were fed forage-based diets and the remaining two pens (3 and 4) were fed a grain-based diet on an *ad libitum* basis, as described in Table S1. Following FP1, the animals in pen 2 were switched from a forage-based to a grain-based diet, and the animals in pen 3 were switched from grain to forage diets (Table S1). Bulls in pens 1 and 4 were exclusively fed forage and grain diets, respectively, in both FPs (Table S1). It is important to mention that a 14-day adaptation period was observed before the commencement of the second feeding period. Nutritional composition of the forage and grain diets is listed in Table S2.

Individual feed intakes were measured using the GrowSafe® feeding system to provide growth and intake data (DMI) needed for the estimation of feed conversion ratio (FCR), which was used as a measure of feed efficiency in this study. FCR was determined as a ratio of DMI to average daily gain (ADG) (computed on a biweekly basis) (Montanholi et al. 2010). Starch and neutral detergent fiber (NDF) contents (Table S2.) of the respective diets were used to calculate starch and NDF intakes from the daily DMI (NRC 2016). Finally, enteric CH_4 emissions were measured in four 24-h periods during Period 1, and in three 24-h periods during Period 2 using the Sulphur hexafluoride (SF_6) technique as described by Thompson (2015).

2.2.1.1 Rumen fluid sampling. Two rumen fluid samples were collected on days 0 and 80 in both FP1 and FP2 using a Geishauser oral probe (Duffield et al. 2004). However, the downstream analysis was performed only on the rumen samples taken on day 80 in each period, as a previous statistical analysis (data not shown) revealed a carryover effect on the rumen microbial abundances between feeding periods likely because the washout

period (adaptation phase) was 14 days (Table S1). Approximately 250 mL of rumen fluid was collected in each sampling, snap frozen using liquid nitrogen, and then stored at -80°C awaiting molecular analysis.

2.2.2 Feed chemical analysis

Briefly, the composited and dried feeds were analyzed for DM (dry matter; method 934.01) as described by AOAC (2000). Both Neutral (NDF; with α-amylase and sodium sulphite) and acid (ADF) detergent fiber (Van Soest et al. 1991) were quantified using Ankom Fiber Analyzer (Ankom Technology Corporation, Macedon, NY, USA). Starch was determined by the α-amylase method as described by Hall (2009). A Leco combustion nitrogen (N) analyzer (FP-428N Determinator, Leco Corporation, St Joseph MI, USA) was used to measure N content. Crude protein (CP) was calculated as N × 6.25 (NRC 2016).

2.2.3 VFA analysis

Supernatants from rumen fluid samples were obtained after centrifugation at 3,000 × g for 15 min at 4˚C and mixed with 25% phosphoric acid (4:1; v/v) for the subsequent gas chromatography (GC) analysis. After adding the internal standard to the samples and incubating them at -20˚C overnight, they were centrifuged at 19,000 × g for 5 min at 4˚C and the supernatant was transferred to the GC vials (1.8 mL). Next, 0.8 mL of the sample was combined with 0.2 mL of 25% phosphoric acid and 0.2 mL of internal standard solution. Standards (for acetic, propionic, isobutyric, butyric, isovaleric, valeric, and caproic acids) were prepared by combining 1 mL of standard solution and 0.2 mL of internal standard solution. The GC analysis was performed using the column Stabilwax-DA 30 meter (Restek Corp), the head pressure of 7.5 psi, split vent flow of 20 mL/minute, and injector temperature of 170ºC.

2.2.4 DNA extraction

Total DNA was extracted from the ruminal content samples using the bead beating method as described by Yu and Morrison (2004). Briefly, frozen rumen content was thawed on ice and 1g of sample was added to 15 mL falcon tubes, washed in 4 mL of TN150 buffer (10 mM Tris-HCl [pH 8.0], 150 mM NaCl), and centrifuged at 14,600 × g for 5 min at 4°C. Thereafter, samples were physically disrupted in a BioSpec Mini Beads

beater 8 (BioSpec, Bartlesville, OK, USA) at 4800 rpm for 3 min, and subjected to a phenol/chloroform/isoamyl alcohol (25: 24: 1) extraction protocol. The DNA was precipitated with cold ethanol and dissolved in nuclease-free water (30 µl). Lastly, the concentration and quality of DNA were measured using Nanodrop® ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA, US). Quantitative real-time PCR (qPCR) was performed only on DNA samples with a ratio of absorbance at 260nm to 280nm higher than 1.8 and a ratio of 260 nm wavelength absorbance to 230 nm between 2.0 and 2.2.

2.2.5 Quantitative real-time PCR analysis

qPCR was performed using SYBR Green chemistry (Fast SYBR® Green Master Mix; Applied Biosystems) on a StepOnePlusTM Real-Time PCR System (Applied Biosystems). The partial bacterial and archaeal 16S rRNA genes (V3-V4 regions) were amplified using U2F/U2R (5'-ACTCCTACGGGAGGCAG-3'; 5'-GACTACCAGGGTATCTAATCC-3') (Stevenson and Weimer 2007) and uniMet1-F/uniMet1-R primer pairs (5'-CCGGAGATGGAACCTGAGAC-3'; 5'-CGGTCTTGCCCAGCTCTTATTC-3') (Zhou et al. 2009). Protozoa and fungi were amplified using P-SSU-316F/P-SSU-539R (5'-GCTTTCGWTGGTAGTGTATT-3'; 5'-CTTGCCCTCYAATCGT WCT-3') (Romero-Perez et al. 2014) and Fungi-F1/Fungi-R1 (5'-GAGGAAGTAAAAGTCGTAACAAGGTTTC-3'; 5'-CAAATTCACAAAGGGTAGGATGATT-3') (Denman and McSweeney 2006) to target 18S rRNA genes and internal transcribed spacer (ITS), respectively. qPCR experiments were performed using the following program: 95 °C for 10 min, followed by 40 cycles of 95 °C for 20 s and 62 °C for 1 min for bacteria, and 95 °C for 20 s, followed by 40 cycles of 95 °C for 3 s and 60 °C for 30s for archaea, protozoa and fungi. A standard curve was constructed using serial dilutions of plasmid DNA containing the 16S rRNA gene sequence for bacteria and methanogens, ITS for fungi and 18S rRNA gene for protozoa. Copy numbers for each standard curve were calculated based on the following equation: $(NL \times A \times 10^{-9}) / (660 \times n)$, where NL was the Avogadro constant (6.02×10^{23}), A was the molecular weight of DNA molecules (ηg), and n was the length of amplicon (bp) (Malmuthuge et al. 2012). The copy number of 16S rRNA genes for total bacteria, total methanogens, and 18S rRNA gene for protozoa and ITS for fungi per sample was

calculated using the equation of Li et al. (2009): (QM × C × DV) / (S × W), where QM was the quantitative mean of the copy number, C was the DNA concentration of each sample (ηg/μL), DV was the dilution volume of extracted DNA (μL), S was the DNA amount subjected to analysis (ηg), and W was the sample weight subjected to DNA extraction (g). qPCR assay efficiency (E) was determined for each primer by running a set of serial dilutions of the target species along with the unknowns, all in triplicates. A plot of Ct vs. Log DNA dilution allowed the efficiency to be calculated as the antilog of the negative reciprocal of the line slope: $E = (10^{-1/slope - 1}) \times 100$. The data generated from the reactions were used for further analysis only if the qPCR assay efficiency was within the range of $90\% \leq E \leq 105\%$.

2.2.6 Stratification of animals according to shifts in microbial abundances

The total abundance of each microbial group (measured by average copy number without log transformation) was determined for each bull according to the dietary regimen, as described previously (see Table S1 and Fig. S6). Groups of animals were created from the magnitude of change in the microbial abundance based on the microorganisms copy numbers recorded for each bull at the end of Period 1 (baselines; D80) compared with the copy numbers recorded at the end of Period 2 (D180) for each sequence of dietary treatment (Fig. S6). Thus, the formation of the groups was obtained by ranking the animals according to the change in the microbial population in response to each dietary treatment (Fig. S6). Log_2 fold change (fc) in microbial copy numbers higher than +1 represented the *High* group and lower than –1 represented the *Low* group. The log_2fc in the copy numbers between +1 and –1 cut-offs were considered as the *Stable* group (Fig. S6).

2.2.7 Statistical analysis

2.2.7.1 Exploring the effect of the diet on the rumen microbes and phenotypic traits

Principal component analysis (Rohart et al. 2017) was initially used to assess the relationship between rumen microbial abundances (bacteria, protozoa, fungi and archaea) and eight phenotypic traits (DMI, NDF intake, starch intake, CH_4 production, CH_4/kg DMI, FCR, Feeding Time in minutes - DUR, and ADG) according to the dietary regimens

(forage vs. grain). Then, we implemented the *PC-corr* (Principal Component-Correlation) algorithm (Ciucci et al. 2017) to generate discriminative functional modules associated with the diets using p-values of Mann-Whitney tests as evaluators, with each module displaying a correlation structure between features (e.g., microbes, phenotypic traits) and the diets. Features were normalized through *z-score* (centered to have mean 0 and scaled to have standard deviation 1) or *log* (logarithm base 10 plus 1 applied to each data element to avoid problems with 0 values) before building *PC-corr* networks in Cytoscape 3.6.0 (Shannon et al. 2003).

2.2.7.2 Investigating the relationship between the phenotypic traits and the microbial population

Two linear mixed effects models (LMMs) were set up using the package *lme4* ('*lmer*' function) (Bates et al. 2015) to investigate the effect of the phenotypic traits on both the microbial populations (Model 1) and the stratification groups obtained from the shifts in the microbial abundances as described above (Model 2). The rumen microbial abundance was $\log_{10}$ – transformed to meet the assumptions of normality, and the phenotypic traits were scaled (centered to have mean 0 and scaled to have standard deviation 1) to facilitate model convergence when required.

In the Model 1, the following parameters were used:

$$\mathcal{Y}_{ijkl} = \mu + \psi_i + \pi_k + \tau_l + s_{ij} + \varepsilon_{ijkl}$$

Where, $\mathcal{Y}_{ijkl}$ is the rumen microbial abundance for the j^{th} subject (j = 1, 2, …n_i) in the i^{th} sequence (i =1, 2, 3, 4) during k^{th} period (j = 1, 2) receiving treatment l (phenotypic traits measured according to the diets); μ is the overall mean; ψ_i is the sequence effect or carryover effect; π_k is the effect of the k^{th} period (k = 1, 2); τ_l is the effect of the lth treatment; s_{ij} is the effect of the jth subject in the ith sequence (j = 1, 2, …n_i ; i =1, 2, 3, 4); ε_{ijkl} is the within-subject error for the jth subject in the ith sequence and the kth period receiving treatment l. Here, it was assumed that ψ, π, and τ are fixed effects, s_{ij} is random effect with mean zero and variance σ_s^2 (between-subject variance), and ε_{ijkl} is the random error with means zero and variances σ^2 (within-subject variance).

In Model 2, the following parametrization was implemented:

$$\mathcal{Y}_{ij} = \mu + \mathcal{X}_{ij} + \mathcal{W}_{ij} + \mathcal{X}_{ij} \times \mathcal{W}_{ij} + \alpha_i + \varepsilon_{ij}$$

Where, $\mathcal{Y}_{ij}$ is the phenotypic traits (CH_4 production, CH_4/kg DMI, FCR, DUR, DMI and ADG) for the j^{th} observation in the subject/bull *i*, and *i* = 1, 2, …n_i; μ is the overall mean; $\mathcal{X}_{ij}$ is the fixed effect of starch and NDF intake measured when bulls were fed forage/grain diets; $\mathcal{W}_{ij}$ represents the fixed effect of the groups determined from the magnitude of change in the host microbial population (high, stable and low); $\mathcal{X}_{ij} \times \mathcal{W}_{ij}$ is the interaction term (fixed effect); α_i is the random effect of bull *i*; ε_{ij} is the residual random error. Here, it was assumed that α_i is normally distributed with mean 0 and variance σ_s^2 (between-subject variance), and ε_{ij} has mean 0 and variance σ^2 (within-subject variance).

The best models were selected using a backward stepwise approach based on Akaike information criteria ('*drop1*' function) (Burnham et al. 2002), with significance of fixed effects and their interactions being tested by comparing the models with a likelihood-ratio test (i.e., chi-squared test) (Romain et al. 2009). Modeling assumptions were also checked by visually inspecting residual patterns (Zuur et al. 2009). All the statistical analysis described in this study were performed using R 3.4.2.

2.3 Results

2.3.1 Effect of diet on the rumen microbes and phenotypic traits

Total bacterial abundance ranged from 0.32 ± 0.282 to 5.9 ± 6.80 × 10^{11} copy numbers of 16S rRNA genes/mL rumen fluid (Table 2.1). Total methanogen abundance varied from 2.5 ± 1.38 to 7.5 ± 5.97 × 10^8 copy numbers of 16S rRNA genes/mL rumen fluid, while the abundances of fungi and protozoa ranged from 0.05 ± 0.062 to 1.1 ± 0.89 × 10^5 and from 1.9 ± 3.83 to 10.0 ± 11.2 × 10^7 copy numbers of ITS/18S rRNA genes/mL rumen fluid, respectively (Table 2.1).

Then, it was evaluated if a relationship could be established between the abundances of the four microbial populations and eight phenotypic traits according to the

dietary regimens (forage vs. grain) (Fig. S1). Application of the *PC-Corr* algorithm showed that all four (bacteria, methanogens, protozoa, and fungi) were clearly separated across diets (P <0.01) and emerged as an interconnected module positively correlated to each other and with a higher abundance in cattle fed forage diets (Figure 2.1). *PC-Corr* also differentiated (P <0.01) phenotypic traits (NDF intake, starch intake, ADG, DMI, CH_4/kg DMI, and FCR) depending on the dietary treatments (Figure 2.1). This result showed that grain diets resulted in a higher DMI, starch intake, and ADG than forage diets (Figure 2.1). In the network structure, those three phenotypic traits arose as a module that was negatively correlated with NDF intake, FCR, and CH_4/kg DMI, which were features displaying higher values in cattle fed forage diets (Figure 2.1).

2.3.2 Effect of the phenotypic traits on the microbial populations

2.3.2.1 Response to the intake of DM and starch, and animal performance (ADG)

Our results revealed that the intake of DM and starch affected (P < 0.05) the abundance of bacteria, reducing it by about 0.2 (±0.06) $\log_{10}$ (Figure 2.2). It should be noted that the diets influenced (P < 0.05) DMI, with bulls consuming grain diets exhibiting a greater DMI (9.6 ± 0.19 kg/day) than those fed forage diets (7.4 ± 0.35 kg/day) (Fig. S5). In a similar fashion, the diets affected (P < 0.05) starch intake, with bulls fed grain diets consuming more starch (3.2 ± 0.05 kg/day) than those fed forage diets (1.5 ± 0.10 kg/day) (Fig. S5). It is worth mentioning that DMI influenced (P < 0.05) the host production of CH_4, increasing it by about 8.7 L/day ± 4.52 (Fig. S2). Although starch intake did not influence the host production of CH_4, it affected (P < 0.05) the host production of CH_4/kg DMI, reducing it by about 1.8 L ± 0.34 CH_4/kg DMI (Fig. S2).

As to the animal performance, our findings showed that the increase in ADG influenced (P < 0.05) the abundance of bacteria, lowering it by about 0.1 (±0.05) $\log_{10}$ (Figure 2.2). The diets also affected (P < 0.05) ADG, with bulls fed grain diets gaining more weight (1.6 ± 0.05 kg/day) than bulls fed forage diets (1.2 ± 0.08 kg/day) (Fig S5).

2.3.2.2 Response to NDF intake, CH_4/kg DMI and VFA concentrations

NDF intake and CH_4/kg DMI influenced (P < 0.05) the abundance of bacteria, increasing it by about 0.1 (±0.05) $\log_{10}$ (Figure 2.3). It is important to note that bulls fed forage diets

consumed more ($P < 0.05$) NDF (3.3 ± 0.10 vs. 1.4 ± 0.06 kg/day) and produced more ($P < 0.05$) CH_4/kg DMI (23.0 ± 1.01 vs. 19.3 ± 0.7 L CH_4/ kg DMI) than bulls fed grain diets (Fig. S5). NDF intake also affected ($P < 0.05$) the host production of CH_4 and CH_4/kg DMI, increasing them by about 9.3 ± 4.2 L/day and 1.4 ± 0.37 L CH_4/kg DMI, respectively (Fig. S2). As expected, NDF intake affected ($P < 0.05$) the total VFA and acetic acid concentrations, increasing them by about 9.0 ± 2.67 and 7.0 ± 1.78 µmol/mL, respectively (Fig. S4). However, the increase in the total VFA concentrations reduced ($P < 0.05$) the abundance of bacteria in about 0.02 (±0.00) log_{10} (Fig. S4). Similarly, a reduction trend ($P < 0.1$) of 0.01 (±0.01) log_{10} in the bacterial abundance was detected with the increase in acetic acid concentrations (Fig. S4). Although we did not find any relationship between NDF intake and propionic acid, bacterial abundance reduced in 0.09 (±0.04) log_{10} when rumen concentrations of propionic acid increased (Fig. S4).

2.3.3 Rumen microbial dynamics in response to the interactions of phenotypic traits and the magnitude of change in host individual microbial population

Since the abundances of methanogens and protozoa did not change significantly with the dietary treatments, they were not considered during stratification of animals according to the shift in microbial abundances. Thus, groups of animals were only created according to the magnitude of change in bacterial and fungal abundances to assess individual variability driven by the diets, and to identify feed efficient animals within the created groups (Fig. S6). Our results showed that shifts from forage to grain resulted in a more pronounced animal-to-animal variation than in individuals that began on grain and then switched to the forage diets (See Figs. S6B and C).

The magnitude of change values from baseline (D80) to D180 revealed three distinct groups of animals based on the grouping cutoffs (Low, log_2-fc < -1; Stable, -1 < log_2-fc < 1; and High, log_2-fc > 1), and the data showed that the interaction between the stratification of these groups and NDF intake influenced (P<0.05) FCR (Figure 2.4). Yet, our analysis showed that both NDF and starch intake influenced ($P < 0.05$) FCR, with NDF intake increasing FCR by about 0.3 (± 0.08) kg DMI/kg gain, and starch intake lowering it ($P < 0.05$) by about 0.2 (± 0.08) kg DMI/kg gain (Fig. S3).

Based on the inter-dependence detected between the stratification of those groups and NDF intake, the bulls classified in the high and low groups determined from the change in the bacterial abundance were more ($P < 0.05$) feed efficient (increase in 0.2 ± 0.09 and 0.1 ± 0.20 kg DMI/kg gain, respectively) than bulls ranked in the stable group (increase in 0.6 ± 0.14 kg DMI/kg gain) (Figure 2.4). To confirm these findings, the feeding time was measured in each group and the results indicated that the bulls of the high/low groups spent less ($P < 0.05$) time feeding than those of the stable group (Figs. 2.4B). A trend ($P < 0.1$) towards an improved feed efficiency in bulls belonging to the high/low groups compared to the stable group was detected in animals ranked according to the fungal abundance change (Figure 2.4C). Finally, animals ranked in the high/low groups according to the fungal abundance change spent less time ($P < 0.05$) in the feed bunker than those of the stable group (Figure 2.4D).

2.4 Discussion

Previous studies have demonstrated that microbial community composition and metabolic potentials in the rumen are remarkably different with respect to nutrient utilization, even in animals raised under the same diet and management regimens (Brulc et al. 2009, Z. P. Li et al. 2016). Thus, a better understanding of interactions between the host and its individualized microbiota is crucial for predicting microbial shifts and identifying individuals that are either responsive (positive and negative responders) or resilient to dietary changes (Bashiardes et al. 2018). Despite being widely investigated using molecular-based approaches, most studies examining diet-driven microbial shifts in the rumen have generated results that are usually qualitative (presence or absence of particular microbial taxa) (Belanche et al. 2011) or semi-quantitative (relative abundance of each microbial taxa) (Comtet-Marre et al. 2017), but not quantitative. To overcome such limitations, it was defined the “baseline” of the quantified rumen microbiota (detected by qPCR) in cattle experiencing dietary changes and developed a strategy to identify feed efficient cattle from the dynamic shift of microbial population abundance. It was also explored the potential to use this approach to assess differences in cattle phenotypes (e.g., FCR, CH_4 emissions).

To our knowledge, this is the first study that quantified the four groups of rumen microbes (bacteria, fungi, protozoa, and methanogens) simultaneously and integrated the microbial data obtained *in vivo* with environmental factors (e.g., diet components). Although previous investigations have focused on those microbial groups and their response to dietary supplementation under *in vitro* conditions (Wang et al. 2017), most *in vivo* studies have concentrated efforts on those microbial groups separately (Tajima et al. 2001, Rico et al. 2015). In the current work, it was monitored the shift in microbial population abundance within the same animal in response to the diet, as well as the response of phenotypic traits to the interactions between the groups of microbial population change (high, low, stable) and the dietary components. Our approach is analogous to that used by biomedical researchers in which individuals are grouped according to the magnitude of microbial change to analyze the effect of dietary factors on microbial composition within the groups of individuals (Martínez et al. 2010, Salonen et al. 2014, Tap et al. 2015). Here, our results clearly showed that shifts from forage to grain (Fig. S6B) resulted in a more pronounced animal-to-animal variation in individuals fed the latter diet, suggesting that differences in microbial abundance drive the resulting variation in microbial baselines between and within diets, and not necessarily by the presence or absence of taxa (Wolff et al. 2017).

The decline in the bacterial population density as starch intake increased (Figure 2.2) did not corroborate past studies that suggested that adding more available energy (starch) to the diet in the form of non-fiber carbohydrates usually favors microbial growth (Grubb and Dehority 1975, Hackmann and Firkins 2015). This study speculated that secondary factors (e.g., individual variability in feed passage rate) in addition to the diet composition might have affected the microbial density. Grain diets usually increase feed passage rate through the gastrointestinal tract, which can lower bacterial abundance in the rumen by decreasing the residence time of feed particles and the subsequent bacterial attachment to the digesta (McAllister et al. 1994a). On the other hand, the models predicted an increase in the bacterial abundance as NDF intake increased (Figure 2.3A), suggesting that bacteria and likely fungi acted in concert to stimulate fiber digestion in those animals. It has been reported that fungi can penetrate and physically disrupt the plant cell wall using an appressorium-like structure (Ho et al. 1988) to increase the surface

area for subsequent bacterial colonization (Lee et al. 2000). However, it is worth mentioning that the differences in the bacterial abundance response to the intake of NDF could have been caused by the type of fiber (alfalfa hay, corn silage, alfalfa silage) present in the feed studied and on their variation (Table S2). Although this study did not investigate the influence of those different types of fiber on the microbial abundance, it has been demonstrated that the rate and extent to which fiber components are degraded depend on the microbial accessibility to feed substrates, which is affected by the physical and chemical nature of the forage (Varga and Kolver 1997). Another prerequisite to be considered is the level of NDF in the diet, as it can influence the time of rumination and consequently the microbial colonization of feed particles (NRC 2016). Therefore, variations in the content of other diet components that no NDF can affect the time of rumination and the microbial fermentation of the feed, and consequently the results found in this study.

Although the bacterial density had a significant influence on CH_4 production per kg DMI (Figure 2.3), *PC-Corr* revealed a functional module constituted by the four groups of rumen microbes in cattle fed forage diets (Figure 2.1B), indicating that CH_4 production may depend on the interactions of methanogens (H_2-consumer) with bacteria, protozoa, and fungi (H_2-producers) to lower the partial pressure of H_2 during the ruminal fermentation (Morgavi et al. 2012a, Janssen 2010, Morgavi et al. 2010). Surprisingly, this study did not find any relationship between methanogens/protozoa and starch/NDF intake, suggesting that protozoa-methanogens associations may have served to protect methanogens from being washed out of the rumen, as protozoa pass the rumen at a slower rate than bacteria and fungi (McAllister et al. 1994a). The lack of effect of total methanogen abundance on CH_4 emissions confirms the findings of Zhou et al. (2011) who reported that the total methanogen numbers did not influence CH_4 production, but rather only the abundance of particular species (e.g., *Methanobrevibacter gottschalkii*) were linked to CH_4 output. However, our models predicted that the ability of NDF to increase CH_4 (Fig. S2) may be related to the increase in total bacteria abundance when NDF consumption and CH_4/kg DMI increased (Figure 2.3), indicating that increased bacterial abundance leads to greater H_2 production which is subsequently used for methanogenesis (Janssen 2010, Leahy et al. 2010, Vanwonterghem et al. 2016). It is

important to note that specific genera/species may be critical for feed digestion, even though in low numbers, which may affect the microbial response to the NDF consumption, CH_4/kg DMI, and consequently the host feed efficiency.

The changes at a single taxa level can be challenging to assess, and it may not directly contribute to the overall rumen fermentation outcomes (such as VFA production) due to the fuctional redundancy of the rumen microbiome. It has been reported that among host-adapted microbes, DNA and RNA abundances would be correlated if many microbial genes were not differentially regulated and were transcribed at the same constant rate (Franzosa et al. 2014). Consequently, variation in microbial composition may not correlate with their metagenomic content (Franzosa et al. 2014) since different taxa may have the same functions. Although in the past a direct relationship between the total densities of bacteria and methanogens with feed efficiency traits was not observed (Zhou et al. 2011, Zhou et al. 2010, Zhou et al. 2009), it is possible that this could be due to the comparatively poor understanding of the roles of protozoa and fungi in the rumen. Results of the current study showed that the dynamics of the four groups of microbes is associated with the variation in the VFA production which is directly linked with host energy metabolism, feed efficiency and methane emissions.

The approach used to create groups of animals based on the variation in the abundance of microbial populations revealed that a relationship could be established between feed conversion efficiency and the groups of microbial change (Figure 2.4). This approach may be implemented in livestock systems as it reflects how individual hosts utilize the diet and how each of them responds to the diet based on the microbial shift. First, it was found that an improved FCR was linked to a higher abundance of bacteria (log_2fc in bacteria > 1), and this result is in line with previous studies showing that bacterial growth is associated with a better animal performance when followed by the increase in bacterial N flow and microbial protein synthesis in the rumen (Sniffen and Robinson 1987, NRC 2016). Second, this study advances current knowledge in this field by demonstrating that a lower bacterial abundance can also be associated with feed efficient cattle as observed in bulls exhibiting a lower bacterial abundance (log_2fc in bacteria < -1) (Figure 2.4). These results suggest that those animals utilized efficiently the energy from the diet

in order to optimize microbial protein synthesis (Turnbaugh et al. 2009) despite the decline in their bacterial abundance. In this case, the reduction in bacterial density may have been accompanied by a decrease in richness as shown by the metatranscriptomic analysis of cattle with lower feed efficiency (Fuyong Li and Le Luo Guan 2017, Neves et al. 2017) and a higher dominance in taxonomic composition (e.g., lactate utilizing bacteria) that favors simpler metabolic networks (e.g. acrylate pathway) (40).

Our interest in the stratification of animals via their magnitude of change in baseline microbiota following dietary changes is driven by previous observations of substantial variation in performance across individuals, even when maintained on the same diet (Z. P. Li et al. 2016, Brulc et al. 2009, Zhou et al. 2018). This research speculated that such variation could be due to differences in rumen microbial function and cattle genetic makeup. In this study, we used cattle with extremely similar background (Red and Black Angus) and aimed to find out at what extent the rumen microbiota abundance differed when environmental traits (such as diet components) were taken into consideration. The findings of this study could benefit animal husbandry in two aspects. First, by understanding the dynamics of the ruminal microbial population in response to dietary changes, researchers will be able to design better feeding strategies to improve the rumen function and host performance, since the microbial community plays an important role in the digestion of feedstuffs (Belanche et al. 2012, Fernando et al. 2010). Second, our approach could serve as a inexpensive strategy to quickly assess rumen microbial shifts in cattle populations experiencing dietary changes under field situations (e. g., feedlots), and to identify feed efficient animals since a relationship between microorganisms abundance and FCR was found in this study.

2.5 Conclusions

This study showed that the individual responses of the rumen microbiota to the dietary treatments reflected the interactions between host and the examined phenotypic traits. The key finding is that the dynamics of the rumen microbial population is intimately associated with inter-individual variability in the baseline microbiota, confirming that the host microbiome individuality may play a more pronounced role in gut response than the dietary change itself. It was also found that bacterial abundance may serve as a useful

proxy measurement to predict changes in feed conversion efficiency and CH_4/kg DMI in cattle. Even though this experiment used simple and inexpensive methods (e.g., ranking animals based on their microbial population shift using qPCR), the results obtained here could be used to design better feeding strategies to enhance the rumen function and to identify cattle with improved feed efficiency based on an individualized microbiota-targeted feeding approach. By ranking animals according to their microbial response to the diet, it was showed that individual hosts exhibiting variability in bacterial abundance ($log_2fc < -1$ or > 1) were more efficient in terms of feed conversion ratio than bulls presenting a stable variability ($-1 > log_2fc < 1$) in bacterial abundance.

2.6 Literature cited

AOAC (2000). *Official Methods of Analysis (17th).* Gaithersburg, MD: Association of Official Analytical Chemists.

Bashiardes, S., Godneva, A., Elinav, E., and Segal, E. (2018). Towards utilization of the human genome and microbiome for personalized nutrition. *Current Opinion in Biotechnology* 51**,** 57-63. doi: 10.1016/j.copbio.2017.11.013.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *2015* 67(1)**,** 48. doi: 10.18637/jss.v067.i01.

Belanche, A., Abecia, L., Holtrop, G., Guada, J.A., Castrillo, C., de la Fuente, G., et al. (2011). Study of the effect of presence or absence of protozoa on rumen fermentation and microbial protein contribution to the chyme1. *Journal of Animal Science* 89(12)**,** 4163-4174. doi: 10.2527/jas.2010-3703.

Belanche, A., Doreau, M., Edwards, J.E., Moorby, J.M., Pinloche, E., and Newbold, C.J. (2012). Shifts in the Rumen Microbiota Due to the Type of Carbohydrate and Level of Protein Ingested by Dairy Cattle Are Associated with Changes in Rumen Fermentation. *Journal of Nutrition* 142(9)**,** 1684-1692. doi: 10.3945/jn.112.159574.

Brulc, J.M., Antonopoulos, D.A., Berg Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., et al. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences* 106(6)**,** 1948-1953. doi: 10.1073/pnas.0806191105.

Burnham, K.P., Anderson, D.R., and Burnham, K.P. (2002). *Model selection and multimodel inference : a practical information-theoretic approach.* New York : Springer, c2002.

2nd ed.

Ciucci, S., Ge, Y., Durán, C., Palladini, A., Martínez-Sánchez, L.M., Cannistraci, C.V., et al. (2017). Enlightening discriminative network functional modules behind principal component analysis separation in differential-omic science studies. *Scientific Reports* 7. doi: 10.1038/srep43946.

Comtet-Marre, S., Parisot, N., Lepercq, P., Chaucheyras-Durand, F., Mosoni, P., Peyretaillade, E., et al. (2017). Metatranscriptomics Reveals the Active Bacterial

and Eukaryotic Fibrolytic Communities in the Rumen of Dairy Cow Fed a Mixed Diet. *Frontiers in Microbiology* 8(67). doi: 10.3389/fmicb.2017.00067.

Denman, S.E., and McSweeney, C.S. (2006). Development of a real-time PCR assay for monitoring anaerobic fungal and cellulolytic bacterial populations within the rumen. *Fems Microbiology Ecology* 58(3)**,** 572-582. doi: 10.1111/j.1574-6941.2006.00190.x.

Duffield, T., Plaizier, J.C., Fairfield, A., Bagg, R., Vessie, G., Dick, P., et al. (2004). Comparison of techniques for measurement of rumen pH in lactating dairy cows. *Journal of Dairy Science* 87(1)**,** 59-66.

Fernando, S.C., Purvis, H.T., II, Najar, F.Z., Sukharnikov, L.O., Krehbiel, C.R., Nagaraja, T.G., et al. (2010). Rumen Microbial Population Dynamics during Adaptation to a High-Grain Diet. *Applied and Environmental Microbiology* 76(22)**,** 7482-7490. doi: 10.1128/aem.00388-10.

Franzosa, E.A., Morgan, X.C., Segata, N., Waldron, L., Reyes, J., Earl, A.M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences* 111(22)**,** E2329-E2338. doi: 10.1073/pnas.1319284111.

Grubb, J.A., and Dehority, B.A. (1975). Effects of an abrupt change in ration from all roughage to high concentrate upon rumen microbial numbers in sheep. *Applied Microbiology* 30(3)**,** 404-412.

Hackmann, T.J., and Firkins, J.L. (2015). Maximizing efficiency of rumen microbial protein production. *Frontiers in Microbiology* 6. doi: 10.3389/fmicb.2015.00465.

Hall, M.B. (2009). Determination of Starch, Including Maltooligosaccharides, in Animal Feeds: Comparison of Methods and a Method Recommended for AOAC Collaborative Study. *Journal of Aoac International* 92(1)**,** 42-49.

Healey, G.R., Murphy, R., Brough, L., Butts, C.A., and Coad, J. (2017). Interindividual variability in gut microbiota and host response to dietary interventions. *Nutrition Reviews* 75(12)**,** 1059-1080.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P.H., et al. (2016). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports* 6. doi: 10.1038/srep19175.

Hernandez-Sanabria, E., Goonewardene, L.A., Wang, Z., Durunna, O.N., Moore, S.S., and Guan, L.L. (2012). Impact of Feed Efficiency and Diet on Adaptive Variations in the Bacterial Community in the Rumen Fluid of Cattle. *Applied and Environmental Microbiology* 78(4)**,** 1203-1214. doi: 10.1128/aem.05114-11.

Ho, Y.W., Abdullah, N., and Jalaludin, S. (1988). Penetrating structures of anaerobic rumen fungi in cattle and swamp buffalo. *Journal of General Microbiology* 134(1)**,** 177-182.

Huws, S.A., Creevey, C.J., Oyama, L.B., Mizrahi, I., Denman, S.E., Popova, M., et al. (2018). Addressing Global Ruminant Agricultural Challenges Through Understanding the Rumen Microbiome: Past, Present, and Future. *Frontiers in Microbiology* 9(2161). doi: 10.3389/fmicb.2018.02161.

IPCC (2006). *Guidelines for National Greenhouse Gas Inventories.* Japan: Institute for Global Environmental Strategies (IGES).

Janssen, P.H. (2010). Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. *Animal Feed Science and Technology* 160(1-2), 1-22. doi: 10.1016/j.anifeedsci.2010.07.002.

Johnson, K.A., and Johnson, D.E. (1995). Methane emissions from cattle. *Journal of Animal Science* 73(8), 2483-2492.

Leahy, S.C., Kelly, W.J., Altermann, E., Ronimus, R.S., Yeoman, C.J., Pacheco, D.M., et al. (2010). The Genome Sequence of the Rumen Methanogen Methanobrevibacter ruminantium Reveals New Possibilities for Controlling Ruminant Methane Emissions. *Plos One* 5(1). doi: 10.1371/journal.pone.0008926.

Lee, S.S., Ha, J.K., and Cheng, K.J. (2000). Relative contributions of bacteria, protozoa, and fungi to in vitro degradation of orchard grass cell walls and their interactions. *Applied and Environmental Microbiology* 66(9), 3807-3813. doi: 10.1128/aem.66.9.3807-3813.2000.

Li, F., and Guan, L.L. (2017). Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Applied And Environmental Microbiology*. doi: 10.1128/AEM.00061-17.

Li, M., Penner, G.B., Hernandez-Sanabria, E., Oba, M., and Guan, L.L. (2009). Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen. *Journal of Applied Microbiology* 107(6), 1924-1934. doi: 10.1111/j.1365-2672.2009.04376.x.

Li, Z.P., Wright, A.D.G., Si, H.Z., Wang, X.X., Qian, W.X., Zhang, Z.G., et al. (2016). Changes in the rumen microbiome and metabolites reveal the effect of host genetics on hybrid crosses. *Environmental Microbiology Reports* 8(6), 1016-1023. doi: 10.1111/1758-2229.12482.

Malmuthuge, N., Li, M., Chen, Y., Fries, P., Griebel, P.J., Baurhoo, B., et al. (2012). Distinct commensal bacteria associated with ingesta and mucosal epithelium in the gastrointestinal tracts of calves and chickens. *Fems Microbiology Ecology* 79(2), 337-347. doi: 10.1111/j.1574-6941.2011.01220.x.

Martínez, I., Jaehyoung, K., Duffy, P.R., Schlegel, V.L., and Walter, J. (2010). Resistant Starches Types 2 and 4 Have Differential Effects on the Composition of the Fecal Microbiota in Human Subjects. *PLoS ONE* 5(11), 1-11. doi: 10.1371/journal.pone.0015046.

McAllister, T.A., Bae, H.D., Jones, G.A., and Cheng, K.J. (1994). Microbial attachment and feed digestion in the rumen. *Journal of Animal Science* 72(11), 3004-3018.

Montanholi, Y.R., Swanson, K.C., Palme, R., Schenkel, F.S., McBride, B.W., Lu, D., et al. (2010). Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. *Animal* 4(5), 692-701. doi: 10.1017/s1751731109991522.

Morgavi, D.P., Forano, E., Martin, C., and Newbold, C.J. (2010). Microbial ecosystem and methanogenesis in ruminants. *Animal* 4(7), 1024-1036. doi: 10.1017/s1751731110000546.

Morgavi, D.P., Kelly, W.J., Janssen, P.H., and Attwood, G.T. (2012). Rumen microbial (meta)genomics and its application to ruminant production. *animal* 7(s1), 184-201. doi: 10.1017/S1751731112000419.

Neves, A.L.A., Li, F., Ghoshal, B., McAllister, T., and Guan, L.L. (2017). Enhancing the Resolution of Rumen Microbial Classification from Metatranscriptomic Data Using Kraken and Mothur. *Frontiers in Microbiology* 8(2445). doi: 10.3389/fmicb.2017.02445.

NRC (2016). *Nutrient Requirements of Beef Cattle.* Washington, D.C., USA: National Academy Press.

Olfert, E., Cross, B., and McWilliams, A. (1993). *Guide to the Care and Use of Experimental Steers.* Ottawa, ON, Canada: Canadian Council on Animal Care.

Rico, D.E., Preston, S.H., Risser, J.M., and Harvatine, K.J. (2015). Rapid changes in key ruminal microbial populations during the induction of and recovery from diet-induced milk fat depression in dairy cows. *British Journal of Nutrition* 114(3)**,** 358-367. doi: 10.1017/s0007114515001865.

Roehe, R., Dewhurst, R.J., Duthie, C.-A., Rooke, J.A., McKain, N., Ross, D.W., et al. (2016). Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. *Plos Genetics* 12(2). doi: 10.1371/journal.pgen.1005846.

Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology* 13(11)**,** 1-19. doi: 10.1371/journal.pcbi.1005752.

Romain, G., Thomas, T., and Margaret, E.K.E. (2009). Ecological Conditions Affect Evolutionary Trajectory in a Predator-Prey System. *Evolution* (3)**,** 641. doi: 10.1111/j.1558-5646.2008.00559.x.

Romero-Perez, A., Okine, E.K., McGinn, S.M., Guan, L.L., Oba, M., Duval, S.M., et al. (2014). The potential of 3-nitrooxypropanol to lower enteric methane emissions from beef cattle. *Journal of Animal Science* 92(10)**,** 4682-4693. doi: 10.2527/jas.2014-7573.

Russell, J.B., and Rychlik, J.L. (2001). Factors that alter rumen microbial ecology. *Science* 292(5519)**,** 1119-1122. doi: 10.1126/science.1058830.

Salonen, A., Lahti, L., Salojärvi, J., Holtrop, G., Korpela, K., Duncan, S.H., et al. (2014). Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *ISME Journal: Multidisciplinary Journal of Microbial Ecology* 8(11)**,** 2218-2230. doi: 10.1038/ismej.2014.63.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11)**,** 2498-2504. doi: 10.1101/gr.1239303.

Sniffen, C.J., and Robinson, P.H. (1987). Microbial growth and flow as influenced by dietary manipulations. *Journal of Dairy Science* 70(2)**,** 425-441.

Stevenson, D.M., and Weimer, P.J. (2007). Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Applied Microbiology and Biotechnology* 75(1)**,** 165-174. doi: 10.1007/s00253-006-0802-y.

Tajima, K., Aminov, R.I., Nagamine, T., Matsui, H., Nakamura, M., and Benno, Y. (2001). Diet-dependent shifts in the bacterial population of the rumen revealed with real-

time PCR. *Applied and Environmental Microbiology* 67(6), 2766-2774. doi: 10.1128/aem.67.6.2766-2774.2001.

Tap, J., Furet, J.-P., Bensaada, M., Philippe, C., Roth, H., Rabot, S., et al. (2015). Gut microbiota richness promotes its stability upon increased dietary fibre intake in healthy adults. *Environmental Microbiology* 17(12), 4954-4964. doi: 10.1111/1462-2920.13006.

Thompson, S. (2015). *The effect of diet type on residual feed intake and the use of infrared thermography as a method to predict efficiency in beef bulls.* M.Sc., University of Manitoba.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., et al. (2009). A core gut microbiome in obese and lean twins. *Nature* 457(7228), 480-484. doi: http://www.nature.com/nature/journal/v457/n7228/suppinfo/nature07540_S1.html .

Van Soest, P.J., Robertson, J.B., and Lewis, B.A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *Journal of Dairy Science* 74(10), 3583-3597.

Vanwonterghem, I., Evans, P.N., Parks, D.H., Jensen, P.D., Woodcroft, B.J., Hugenholtz, P., et al. (2016). Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nature Microbiology* 1, 16170. doi: 10.1038/nmicrobiol.2016.170

http://www.nature.com/articles/nmicrobiol2016170 - supplementary-information.

Varga, G.A., and Kolver, E.S. (1997). Microbial and animal limitations to fiber digestion and utilization. *The Journal Of Nutrition* 127(5 Suppl), 819S-823S. doi: 10.1093/jn/127.5.819S.

Wang, S., Giller, K., Kreuzer, M., Ulbrich, S.E., Braun, U., and Schwarm, A. (2017). Contribution of Ruminal Fungi, Archaea, Protozoa, and Bacteria to the Methane Suppression Caused by Oilseed Supplemented Diets. *Frontiers in Microbiology* 8(1864). doi: 10.3389/fmicb.2017.01864.

Weimer, P.J. (2015). Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. *Frontiers in Microbiology* 6(296). doi: 10.3389/fmicb.2015.00296.

Wolff, S.M., Ellison, M.J., Hao, Y., Cockrum, R.R., Austin, K.J., Baraboo, M., et al. (2017). Diet shifts provoke complex and variable changes in the metabolic networks of the ruminal microbiome. *Microbiome* 5(1), 60. doi: 10.1186/s40168-017-0274-6.

Yang, T., Wei, Z., Friman, V.-P., Xu, Y., Shen, Q., Kowalchuk, G.A., et al. (2017). Resource availability modulates biodiversity-invasion relationships by altering competitive interactions. *Environmental Microbiology* 19(8), 2984-2991. doi: 10.1111/1462-2920.13708.

Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *BioTechniques* 36(5), 808-812. doi: 10.2144/04365ST04.

Zhou, M., Chung, Y.H., Beauchemin, K.A., Holtshausen, L., Oba, M., McAllister, T.A., et al. (2011). Relationship between rumen methanogens and methane production in dairy cows fed diets supplemented with a feed enzyme additive. *Journal of Applied Microbiology* 111(5), 1148-1158. doi: 10.1111/j.1365-2672.2011.05126.x.

Zhou, M., Hernandez-Sanabria, E., and Guan, L.L. (2009). Assessment of the Microbial Ecology of Ruminal Methanogens in Cattle with Different Feed Efficiencies. *Applied and Environmental Microbiology* 75(20), 6524-6533. doi: 10.1128/aem.02815-08.

Zhou, M., Hernandez-Sanabria, E., and Guan, L.L. (2010). Characterization of Variation in Rumen Methanogenic Communities under Different Dietary and Host Feed Efficiency Conditions, as Determined by PCR-Denaturing Gradient Gel Electrophoresis Analysis. *Applied and Environmental Microbiology* 76(12), 3776-3786. doi: 10.1128/aem.00010-10.

Zhou, M., Hünerberg, M., Beauchemin, K.A., McAllister, T.A., Okine, E.K., and Guan, L.L. (2012). Individuality of ruminal methanogen/protozoa populations in beef cattle fed diets containing dried distillers' grain with solubles. *Acta Agriculturae Scandinavica: Section A, Animal Science* 62(4), 273-288. doi: 10.1080/09064702.2013.788206.

Zhou, M., Peng, Y.-J., Chen, Y., Klinger, C.M., Oba, M., Liu, J.-X., et al. (2018). Assessment of microbiome changes after rumen transfaunation: implications on improving feed efficiency in beef cattle. *Microbiome* 6(1), 62-62. doi: 10.1186/s40168-018-0447-y.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. (2009). "Mixed Effects Models and Extensions in Ecology with R," in *Mixed Effects Models and Extensions in Ecology with R*.).

2.7 Tables and figures

Table 2.1 Quantification of the copy numbers of microbial populations in the rumen of beef cattle fed forage or grain diets

Microorganisms	Sequence of dietary treatments[4]							
	Forage - Forage	Forage - Grain	Grain - Forage	Grain - Grain	Forage - Forage	Forage - Grain	Grain - Forage	Grain - Grain
Bacteria[1]	5.9 ± 6.80 × 10^{11}	2.5 ± 1.67 × 10^{11}	6.0 ± 4.24 × 10^{10}	3.2 ± 2.82 × 10^{10}	5.8 ± 4.39 × 10^{10}	1.9 ± 1.46 × 10^{11}	5.0 ± 3.08 × 10^{11}	5.7 ± 3.44 × 10^{11}
Fungi[2]	1.2 ± 1.03 × 10^{4}	9.1± 7.06 × 10^{3}	1.1 ± 0.89 × 10^{5}	8.8 ± 8.6 × 10^{4}	5.0 ± 6.28 × 10^{3}	2.4 ± 4.59 × 10^{4}	2.1 ± 3.89 × 10^{4}	2.6 ± 2.79 × 10^{4}
Methanogens[1]	3.7 ± 1.71 × 10^{8}	3.1± 0.93 × 10^{8}	2.9 ± 1.29 × 10^{8}	2.5 ± 1.38 × 10^{8}	3.0 ± 1.68 × 10^{8}	4.4 ± 2.91 × 10^{8}	5.4 ± 3.95 × 10^{8}	7.5 ± 5.97 × 10^{8}
Protozoa[3]	4.8 ± 7.73 × 10^{7}	2.9 ± 4.10 × 10^{7}	7.6 ± 0.10 × 10^{7}	6.4 ± 0.11 × 10^{7}	2.6 ± 3.06 × 10^{7}	1.9 ± 3.83 × 10^{7}	1.0 ± 1.12 × 10^{8}	6.8 ± 6.14 × 10^{7}

[1]Copy number of 16S rRNA (Mean ± SD)／mL of rumen fluid;

[2]Copy number of ITS (Mean ± SD)／mL of rumen fluid;

[3]Copy number of 18S rRNA (Mean ± SD)／mL of rumen fluid;

[4]The hyphen (-) means diet change.

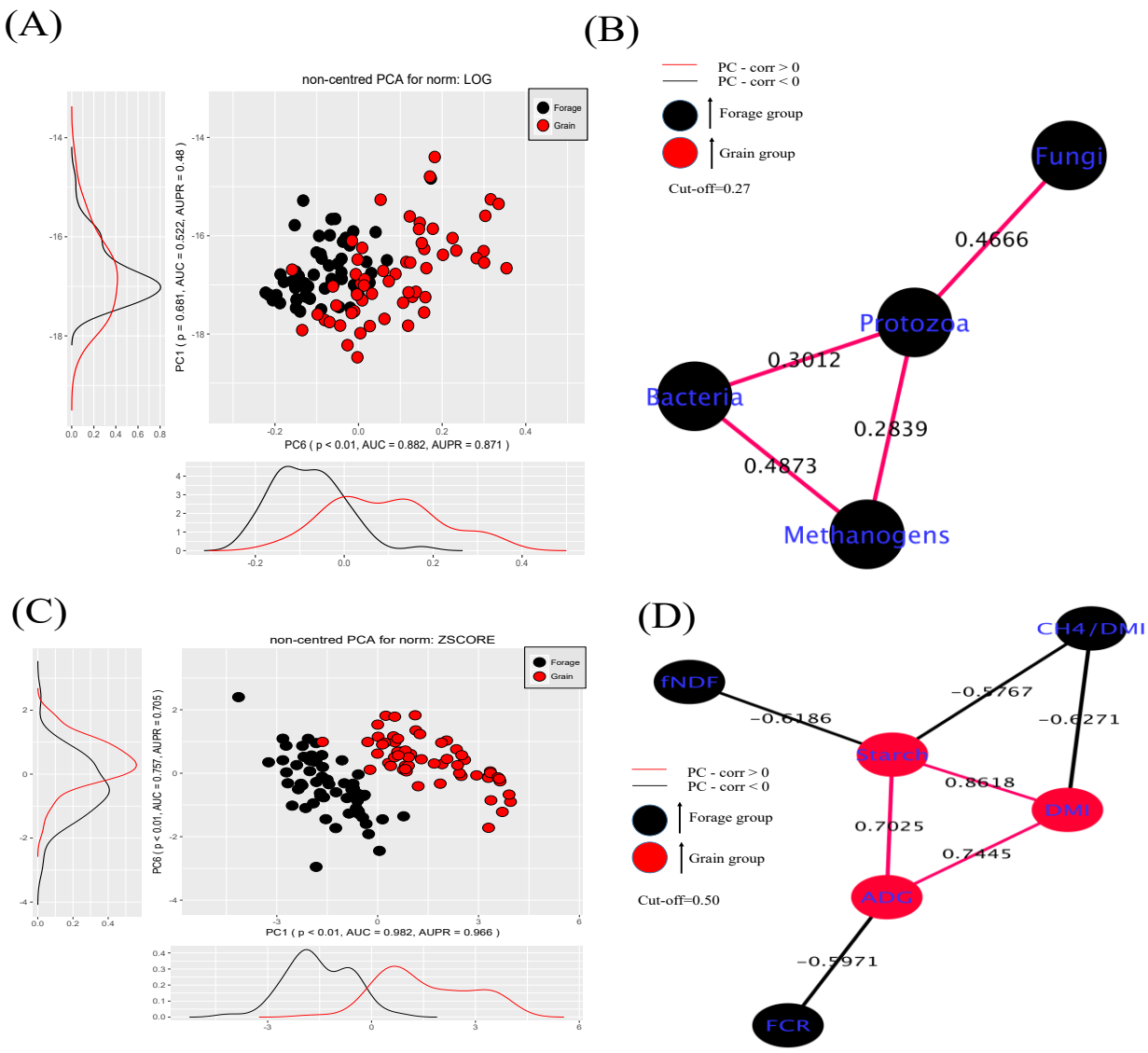


Figure 2.1 Rumen microbial population and phenotypic traits in response to diet changes. Linear dimensionality reduction by *PC-Corr* separated rumen microbes (A) and phenotypic traits (C) recorded in cattle fed forage (black) or grain diets (red). *PC-Corr* detected differences (Mann-Whitney test, P <0.01) in rumen microbes between forage and grain diets along principal component 6 (PC6), whose loadings were used to build a microbial network at a cut-off (Pearson correlation) of 0.27 (B). Differences (Mann-Whitney test, P <0.01) in phenotypic traits were also observed for forage and grain diets along PC1 and PC6, whose loadings were used for the construction of a network at a cut-off of 0.50 (D).

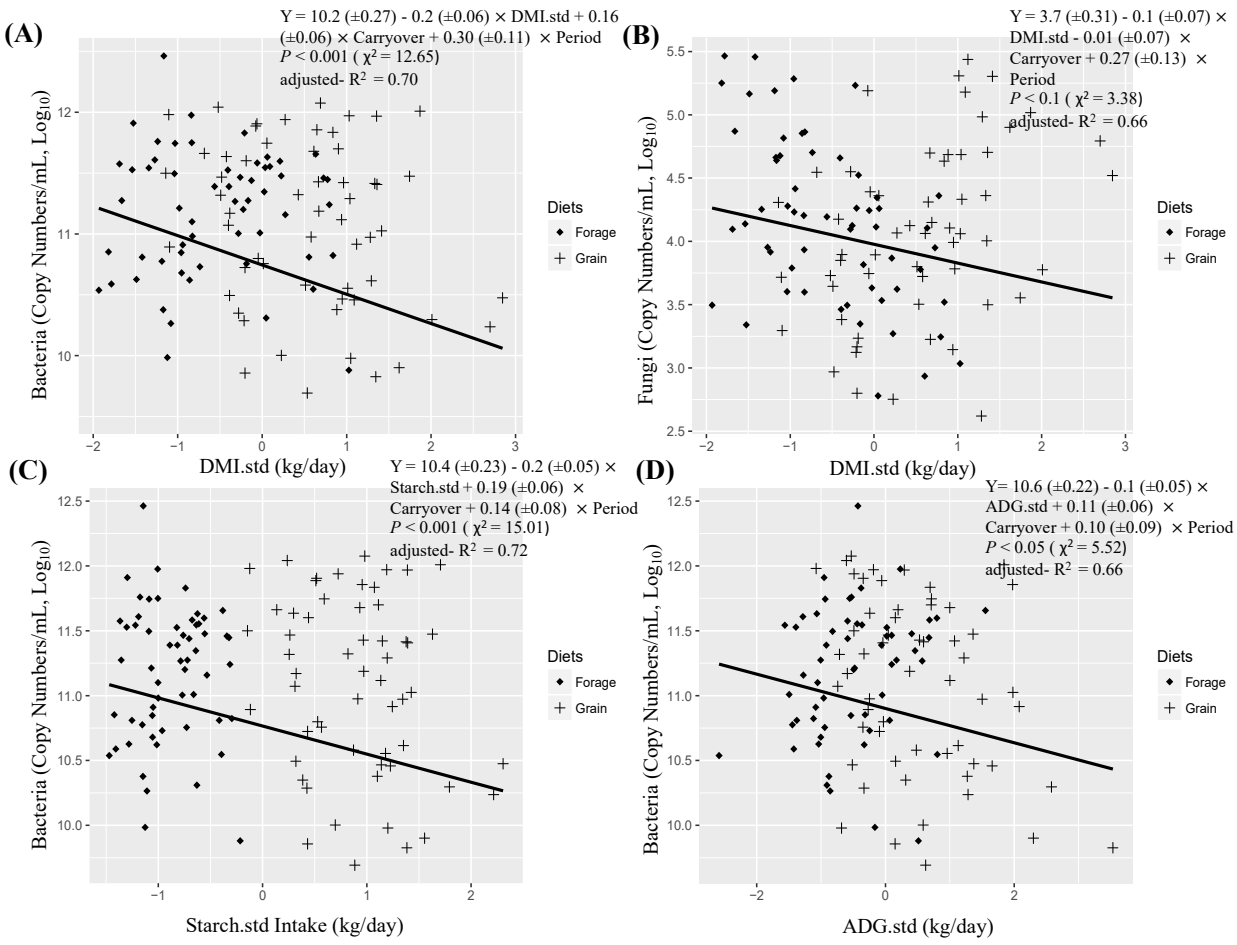


Figure 2.2 Dynamics of rumen bacteria and fungi in response to DMI, starch intake, and average daily gain (ADG). Effect of DMI (A and B), starch intake (C) (kg/day), and ADG (kg/day) on the copy numbers of bacteria and fungi (Log_{10}).

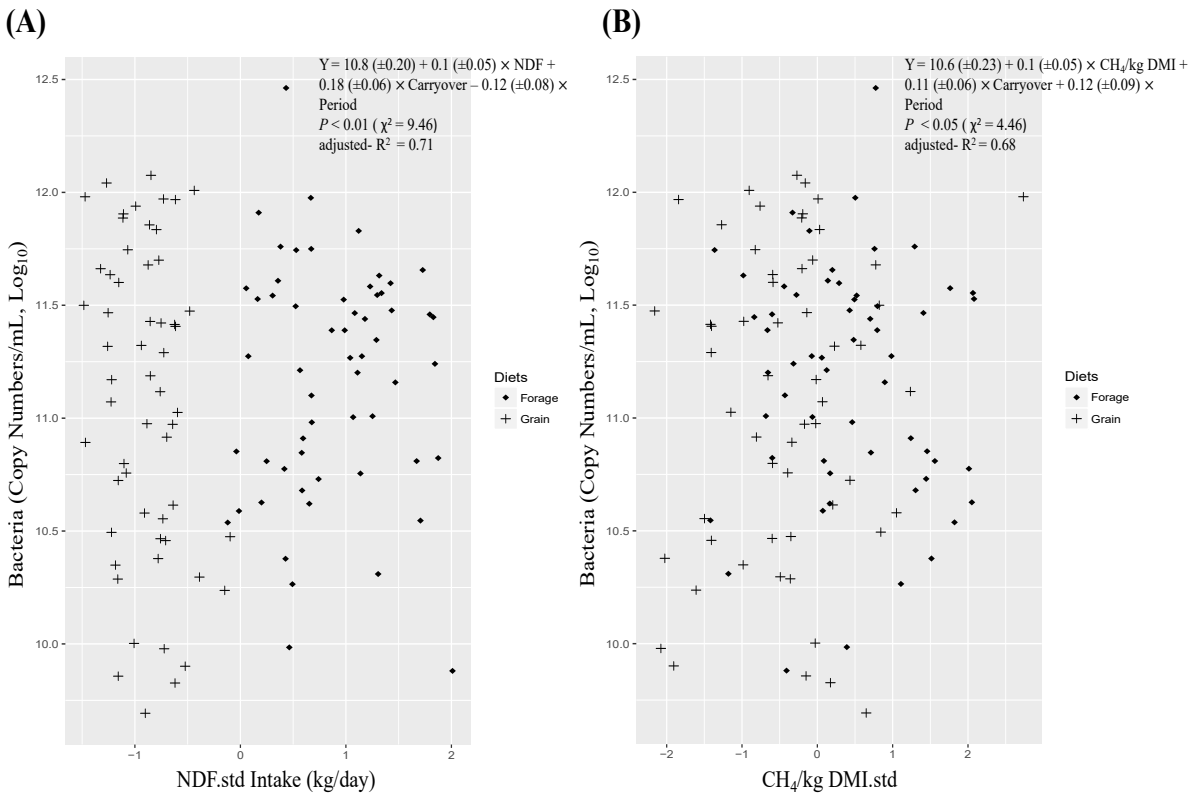


Figure 2.3 Dynamics of rumen bacteria in response to NDF intake and CH_4/kg DMI. Effect of NDF intake (A) and CH_4/kg DMI (B) (kg/day) on the copy numbers of bacteria (Log_{10}).

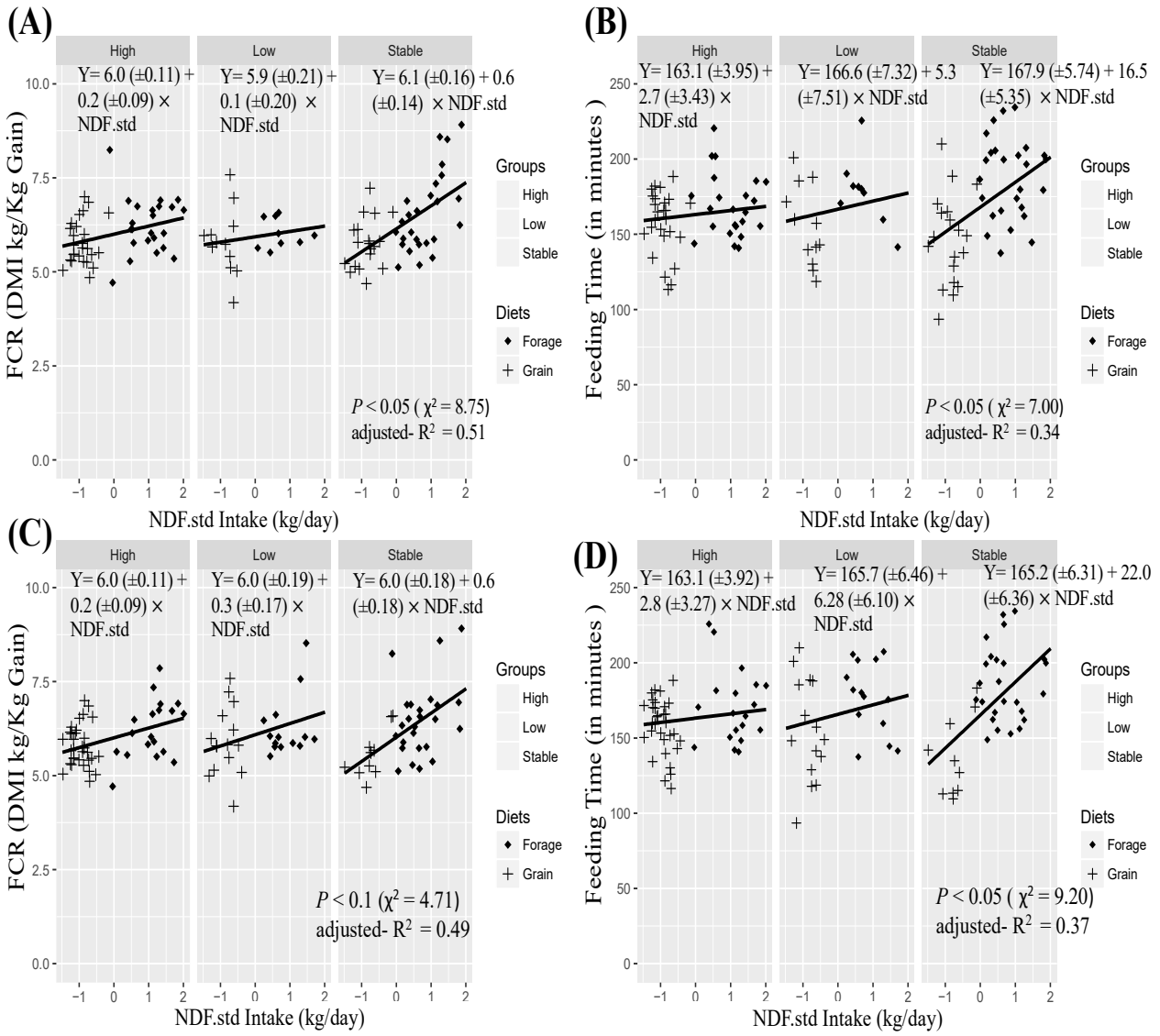


Figure 2.4 Feed conversion ratio (FCR) and Feeding Time (in minutes) in response to the interactions of NDF intake and the magnitude of change in host microbial population. Interactive effects of NDF intake (kg/day) and the magnitude of microbial shift (High, Stable, Low) on FCR and Feeding Time when groups of bulls were created from the magnitude of change in bacteria (A and B) and fungi (C and D).

Chapter 3

Enhancing the resolution of rumen microbial classification from metatranscriptomic data using Kraken and Mothur

3.1 Introduction

The success of microbiome studies (composition, structure, diversity, and function) is primarily ascribable to the development of bioinformatics tools specially tailored to overcome the technical challenges posed by the analysis of massively paralleled, high-throughput sequencing data (Simon and Daniel, 2011; Siegwald et al., 2017). These bioinformatics tools make use of several techniques (e.g., read mapping, k-mer alignment, and composition analysis) (Piro et al., 2017) and can be categorized into two distinct groups: 1) programs that use all available genome sequences (Lindgreen et al., 2016), also called assignment-first approaches (Siegwald et al., 2017) (e.g., CLARK - Ounit et al., 2015; GOTTCHA - Freitas et al., 2015; KRAKEN - Wood and Salzberg, 2014; MG-RAST - Meyer et al., 2008), and 2) programs that target a set of marker genes (Lindgreen et al., 2016), also known as clustering-first approaches (Siegwald et al., 2017) (e.g., QIIME - Caporaso et al., 2010; MOTHUR - Schloss et al., 2009; MetaPhlAn - Segata et al., 2012; mOTU - Sunagawa et al., 2013). In the assignment-first tools, all reads are assigned to the lowest taxonomy unit (lower common ancestor-LCA) within a reference database based on their annotations, while in the clustering-first approaches the reads are grouped into Operational Taxonomic Units (OTUs) using different OTU picking strategies (closed or open reference) to assign reads to a taxonomic group based on their sequence similarities (Siegwald et al., 2017).

However, most of the above studies are focused on demonstrating how single analytical steps (e.g., sequence pre-processing, OTU clustering or taxonomic assignment) generated by the existing tools impact the microbial classification in real or simulated datasets derived from the Human Microbiome Project (Siegwald et al., 2017). Comparison of methodologies to comprehensively classify the rumen microbiome is lacking which may be in part due to its complexity, as the rumen microbial community consists of bacteria, archaea, protozoa and fungi (Russell and Rychlik, 2001). A recent study by Li et al. (2016a) developed a Mothur (Schloss et al., 2009) based pipeline to assess active rumen microbiota from data generated from total RNA sequencing. Later, the same researchers applied this pipeline to

investigate linkages between the active rumen microbiome (structure and function) and feed efficiency in beef cattle using metatranscriptomics (Li and Guan, 2017). Using the developed mothur-based pipeline for taxonomic assignment, the authors identified that the active microbial taxa differed in the rumen of cattle with differing feed efficiency and suggested that the active rumen microbiome is one of the biological factors that may contribute to variations in feed efficiency of beef cattle (Li and Guan, 2017). There were two steps employed in taxonomic classification by Li et al. (2016a): bacterial sequences belonging to V1-V3 regions were extracted from the aligned Greengenes database, and archaeal sequences belonging to the V6-V8 regions were aligned with a rumen-specific archaeal 16S rRNA gene database (Janssen and Kirs, 2008). Despite the efficacy of this pipeline, it still remains a challenge for researchers to determine which approach (assignment- or clustering-first methods) of taxonomic classification delivers the most realistic representation of rumen microbial ecology.

In the current study, we propose a comparative analysis of the outcomes of Kraken (Wood and Salzberg, 2014) and the pipeline of Li et al. (2016a) with a focus on the biological interpretation of the rumen microbial classification from the perspective of two conceptually different software packages. Unlike the pipeline developed by Li et al. (2016a), Kraken algorithms can make multiple comparisons of single or assembled k-mers against any hypervariable region, providing useful information regarding a particular species detected in a region of the 16S rRNA gene that is different from the targeted internal conserved region initially sequenced (Wood and Salzberg, 2014; Valenzuela-González et al., 2016). Although Kraken algorithms have been originally designed to assign taxonomic identity to short DNA reads (Wood and Salzberg, 2014), studies have shown that Kraken is also useful to provide taxonomic classification for long (up to 1352.1 ± 153.72 bp) metagenomic DNA sequences (Valenzuela-González et al., 2016). Therefore, the objectives of this study were (i) to compare and contrast the pipeline of Li et al. (2016a) and Kraken to assess the taxonomic profiles of rumen bacteria and archaea and (ii) to investigate the impact of the comparative analysis of both analytical approaches on the biological interpretation of the rumen microbial classification obtained from cattle exhibiting different feed efficiencies.

3.2 Materials and methods

3.2.1 Animal study and sampling

The experimental procedures described in this study were approved by the Veterinary Services and the Animal Care Committee, University of Manitoba, Canada, to ensure that animals were cared for in compliance with those ethics. Rumen contents were collected from 12 purebred Angus bulls (mean age of 249 ± 22 days and average body weight of 313.9 ± 32 kg) raised in confinement at the Glenlea Research Station located at the University of Manitoba according to the guidelines of the Canadian Council on Animal Care (CCAC) (Olfert et al., 1993), with bulls being fed a forage diet over two 80-day feeding periods (with a 20-day adaptation in between) as described by Thompson (2015). In the current study, 250 ml of rumen contents (liquid and solid fractions) were collected at the end of the second feeding period using a Geishauser oral probe (Duffield et al., 2004), immediately snap frozen in liquid nitrogen, and stored at -80°C for later processing. The feed intake of individual bulls was recorded using the GrowSafe® feeding system (GrowSafe Systems Ltd., Airdrie, Alberta, CA) and the feed conversion rate (FCR) was calculated as a ratio of dry matter intake to average daily gain (computed on a biweekly basis; Montanholi et al., 2010). The bulls were ranked into two groups: high (n=6) and low (n=6) FCR, with high (H-FCR) and low (L-FCR) standing for inefficient and efficient cattle in terms of diet utilization, respectively.

3.2.2 RNA extraction and sequencing

Total RNA was extracted from rumen samples using the TRIzol protocol based on the acid guanidinium-phenol-chloroform method (Chomczynski and Sacchi, 2006; Béra-Maillet et al., 2009) with the modified procedures described by Li et al. (2016a). Briefly, ~200 mg of rumen sample was subjected to RNA extraction with the addition of 1.5 ml of TRIzol reagent (Invitrogen, Carlsbad, CA, USA), followed by 0.4 ml of chloroform, 0.3 ml of isopropanol, and 0.3 ml of high salt solution (1.2 M sodium acetate, 0.8 M NaCl) for the extraction protocol (Li et al., 2016a). The yield and integrity of the RNA samples were determined using a Qubit 2.0 fluorimeter (Invitrogen, Carlsbad, CA, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA samples were subjected to downstream RNA-sequencing only if they exhibited RNA with integrity number (RIN) higher than 7.0. Briefly, total RNA (100 ng) of each sample was used for library construction using the TruSeq RNA sample prep v2 LS kit (Illumina, San Diego, CA, USA) without the mRNA enrichment step (Li et al., 2016a). The quality of libraries was assessed using Agilent 2200 TapeStation (Agilent Technologies) and Qubit 2.0 fluorimeter (Invitrogen). Finally, cDNA fragments (~140 bp) were

paired-end (2 X 100 bp) sequenced using an Illumina HiSeq 2000 system at the McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada).

3.2.3 Pipeline settings

A flow chart is shown in Figure 3.1 to present the software parameters used to obtain the microbial classification from either Mothur (Schloss et al., 2009) or Kraken (Wood and Salzberg, 2014) taxonomic assignment strategies. In the pre-processing steps, all fastq-formatted sequences were firstly uploaded into FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ for quality control and removal of ambiguous sequences, and then the software Trimmomatic (version 0.32; Bolger et al., 2014) was used to trim residual artificial sequences, cut bases with quality scores below 20, and remove reads shorter than 50 bp (Li et al., 2016a). After pre-processing, SortMeRNA (version 1.9; Kopylova et al., 2012) was used to sort the filtered reads into fragments of 16S rRNA (for taxonomic identification using Mothur) based on the rRNA reference databases SILVA_SSU (release 119; Quast et al., 2013) and mRNA (for microbial classification using Kraken). In the pipeline developed by Li et al. (2016a), sorted paired-end reads belonging to bacterial and archaeal 16S rRNA were joined to increase the read length by combining the forward and reverse sequences. After the 16S rRNA sequences were enriched, downstream analyses were performed using Mothur (version 1.31.2; Schloss et al., 2009) as described by Kozich et al. (2013) (Figure 3.1). For taxonomic classification, bacterial and archaeal 16S rRNA sequences were aligned with the V1-V3 region-enriched Greengenes database (DeSantis et al., 2006) and the V6-V8 region-enriched rumen-specific archaea database (Janssen and Kirs, 2008, which was updated from Kittelmann et al., 2013), respectively. De novo chimera detection was then conducted using UCHIME (Edgar et al., 2011), and non-chimeric sequences were taxonomically assessed using a naive Bayesian method (Wang et al., 2007). The pipeline developed by Li et al. (2016a) will be referred as Mothur through the rest of the paper.

As for the Kraken pipeline (Wood and Salzberg, 2014), newly developed *Perl* scripts were used to retrieve all complete genomes of bacteria (5,294) and archaea (209) from NCBI (RefSeq) (May 2016), to build a Kraken standard database (June 2016) based on their annotations at the lowest taxonomic level (Figure 3.1). Ninety-one complete genomes from organisms isolated from the rumen or from ruminant feces or saliva deposited in the

Hungate1000 project were also retrieved from JGI's IMG database (using NCBI Taxon IDs). After downloading the genomes, the script *kraken-build* (option *--build*) was used to set the lowest common ancestors (LCAs) in a bacteria-archaea joint database (size: 115G; number of sequences mapped to profiles: 10,174; and time for database construction: 6h33m35s). Thereafter, each pair of mRNA sequences was assembled by MEGAHIT (Li et al., 2015), with the resulting contigs (with average extension of 472.31 ± 31.10 bp) being assigned by Kraken (through k-mer discrimination) to the LCA in the customized standard database for microbial classification (Figure 3.1). Full taxonomic names associated with each classified sequence (separated from unclassified reads using *kraken* option *--preload*) and standard ranks (from domain to species) for each taxon were provided by *kraken-translate* and *kraken-mpa-report* (Figure 3.1).

3.2.4 Statistical analysis

In this study, a phylotype was considered as classified by both methods if it had at least one count detected in the 12 samples. For comparisons between H-FCR and L-FCR groups, we investigated only bacterial and archaeal profiles with a relative abundance > 0.1% prevalent in at least three samples (3 out 6) to avoid sparsely observed counts, which tend to introduce noise in the analysis (Chen and Li, 2013). The ANCOM procedure (Mandal et al., 2015), which uses an alternative normalization approach called Aitchison's log-ratio transformation (Aitchison, 1982), was then used to normalize the sequence data and to compare the normalized log ratio of the abundance of each taxon to the abundance of all remaining taxa (Weiss et al., 2017). To deal with zero counts in the datasets, ANCOM used an arbitrary pseudo count value of 0.001 (Mandal et al., 2015). Thereafter, Wilcoxon rank sum tests were calculated on each log ratio to find differences between feed efficiency groups (H-FCR vs. L-FCR) as provided by each classification method (Mothur or Kraken) (Figure 3.1). The p-value of each test were adjusted into false discovery rate (FDR) using the Benjamini-Hochberg algorithm (Benjamini and Hochberg, 1995), and a threshold of FDR lower than 0.15 (Korpela et al., 2016) was applied to determine the significance due to the small sample size of this study. Correlation circle plots and relevance networks for core bacterial genera and archaeal species (with a relative abundance > 0.1% detected in all rumen samples; Li and Guan, 2017) were generated from the output of regularized canonical correlation (rCC) analysis as implemented in the R package *mixOmics* (Gonzalez et al., 2008) and Cytoscape 3.4.0 (Shannon et al., 2003). Before running rCC analysis, the data was normalized by total sum

scaling (TSS) (dividing each taxon count by the total number of counts in each individual sample to account for uneven sequencing depths across samples) and then transformed by centered log ratio to project the data from a simplex to a Euclidian space (Aitchison, 1982; Mandal et al., 2015; Cao et al., 2016). Then, estimation of regularization parameters (λ1 and λ2) and canonical correlations were calculated using the cross-validation procedure (Gonzalez et al., 2008). Finally, alpha-diversity indexes were calculated using the R package *vegan* (as provided by each classification method) and compared between FCR groups (H-FCR vs. L-FCR) using paired Wilcoxon signed rank test. All statistical procedures were performed using R 3.3.2 (R Core Team, 2016).

3.2.5 Data submission

The datasets analyzed in this study were submitted to NCBI Sequence Read Archive (SRA) under the accession number PRJNA403833.

3.3 Results

3.3.1 Taxonomic distribution of the microbial profiles performed by Mothur or Kraken

In this study, two bioinformatics approaches, Kraken and a Mothur-based pipeline developed in-house by Li et al. (2016a), were used to obtain taxonomic classifications (bacteria and archaea) of the ruminal microbiota in bulls exhibiting different ($P < 0.05$) feed efficiencies (average FCR for H-FCR group= 7.64 kg dry matter intake (DMI)/kg gain; average FCR for L-FCR group= 5.71 kg DMI/kg gain; $P = 0.008$). Taking into consideration the total number of microbial taxa in the samples, Kraken identified a higher number of bacterial and archaeal phylotypes at all taxonomic ranks than Mothur (Table 3.1). At the phylum level, the results of bacterial profiles revealed a similar taxa distribution of the most abundant taxa classified by both methods (Tables 3.1 and 3.2), with Bacteroidetes, Firmicutes, and Proteobacteria being highly abundant and accounting for approximately 80% of the total bacterial community. However, Spirochaetes (4.9%) were the fourth-most abundant taxon identified by Kraken, followed by Verrucomicrobia (2.3%), Actinobacteria (2.1%), Tenericutes (1.9%), and Fibrobacteres (1.2%). In contrast, Fibrobacteres (3.4%) was found to be the fourth-most abundant taxon detected by Mothur, followed by Spirochaetes (2.2%), Verrucomicrobia (1.7%), Tenericutes (0.8%), and Cyanobacteria (0.6%). Although there was some congruency (69 commonly detected taxa) at the most resolvable level (up to genus) of bacteria in between the two pipelines, an additional 159 genera were exclusively identified by Kraken. Genera

such as *Ruminiclostridium, Lachnoclostridium,* and *Acholeplasma* were uniquely identified by Kraken, whereas *Ruminobacter, Coprococcus, YRC22, and Oscillospira* were exclusively detected by Mothur. As for the most abundant genera, Kraken revealed *Prevotella* (33.5%), *Treponema* (4.1%), *Ruminoccocus* (4.1%), *Ruminiclostridium* (3.2%), *Bacteroides* (3.0%), *Butyrivibrio* (2.4%) and *Clostridium* (2.2%) at relatively high abundances, while Mothur identified *Prevotella* (22.6%), *Ruminoccocus* (14.6%), *Ruminobacter* (4.9%), *Fibrobacter* (4.3%), *Treponema* (2.4%), and *Butyrivibrio* (1.2%) as more abundant. It is worth noting that although Mothur could theoretically classify sequences at the species level, it was not able to assign bacterial contigs further than the genus level in the current study. Conversely, Kraken detected 423 species (Tables 3.1 and 3.2) such as *Prevotella ruminicola* (27.6%), *Butyrivibrio proteoclasticus* (2.8%), *Treponema succinifaciens* (2.6%), *Ruminiclostridium* sp KB18 (2.2%), and *Fibrobacter succinogenes* (1.8%). A complete list of all bacteria phylotypes (in all taxonomic ranks) classified by Mothur or Kraken is provided in Supplementary Tables 1 and 2, respectively. In addition, the direct comparisons of the bacterial taxonomic assignments obtained from both methods across all samples are included in Supplementary Table 3.

In terms of archaea identification, both methods exhibited similar results on the abundance of Methanomassiliicoccaceae (previously referred to as RCC), which comprised more than 65% of the total archaeal families (Table 3.3). However, the two methods generated significantly different archaeal profiles at the species level, with 7 species being exclusively identified by Kraken and 4 taxa being exclusively detected by Mothur (Tables 3.1 and 3.3). Only *Methanobrevibacter ruminantium* was commonly detected by the two methods, being the second-most abundant species classified by Mothur and the seventh-most abundant identified by Kraken. A detailed list of archaeal classification (in all taxonomic ranks) for Mothur or Kraken can be found in the Supplementary Tables 1 and 2, respectively, together with the information on the direct comparison of the archaeal taxonomic assignments obtained from both methods across all samples included in Supplementary Table 4.

3.3.2 Differences in relative abundances of taxa in H- vs. L-FCR rumen samples

To evaluate how the above two approaches affect the biological interpretation of bacteria and archaea diversity and community structure, comparisons of rumen microbiota between H- and L-FCR cattle were performed. Differences in microbial abundance between H- and L-FCR datasets were found to be minimal (making up less than 1% of the total microbial

community), regardless of the classification method (Tables 3.2 and 3.3). In this regard, only the family R4-41B (exclusively detected by Mothur) were more (FDR < 0.15) abundant in the rumen of L-FCR bulls, while the family Actinomycetaceae was more (FDR < 0.15) abundant in L-FCR samples classified by Kraken (Table 3.2). Methanococcaceae and *Xenorhabdus* exhibited a higher (FDR < 0.15) abundance in the rumen of H-FCR bulls when sequences were exclusively classified by Kraken (Tables 3.2 and 3.3).

In addition, alpha-diversity indexes of bacteria (genus level) and archaea (species level) were compared between H- and L-FCR groups to determine how the two pipelines differed in microbial biodiversity estimates. Shannon, Inverse Simpson and Simpson (with rarefy) indexes were higher ($P < 0.05$, paired Wilcoxon signed rank test) in H-FCR than in L-FCR bulls as shown by both pipelines (Table 3.4). On the other hand, a higher ($P < 0.05$, paired Wilcoxon signed rank test) archaeal diversity in the H-FCR group was observed only by the Kraken pipeline (Table 3.4).

3.3.3 Potential interactions between bacteria and archaea detected by Mothur or Kraken

To investigate interactions among different taxa classified by Kraken or Mothur, rCC analysis was implemented to identify relationships *within* and *between* bacteria and archaea communities. Our results revealed that bacteria and archaea interactions were quite contrasting between the two methods, with the microbial groups exhibiting different correlation outcomes as shown in Figure 3.2. *Within* bacterial communities, negative correlations between *Prevotella*, *Treponema, Fibrobacter* and *Ruminobacter*, *Butyrivibrio*, and *Ruminoccocus* were observed using the Mothur pipeline (Figure 3.2a), while *Prevotella* and *Bacteroides* were negatively correlated with *Treponema, Fibrobacter* and *Ruminoccocus* when Kraken was used (Figure 3.2c). Associations *within* archaeal species were also different between the two methods, with *Methanobrevibacter gottschalkii* and *Methanobrevibacter ruminantium* being negatively correlated with each other from the Mothur pipeline, and *Candidatus Methanoplasma termitum* and *Candidatus Methanomethylophilus alvus* exhibiting negative correlations with each other in the Kraken pipeline (Figures 3.2a and 3.2c). Relevance networks of the associations *between* bacteria and archaea revealed a positive correlation between *Methanobrevibacter ruminantium* and *Fibrobacter*, *RFN20*, *Treponema,* and *BF311*, and a positive correlation between *Methanobrevibacter gottschalkii* and

Ruminococcus, Butyrivibrio and *Succiniclasticum* based on the microbial classification by Mothur (Figure 3.2b). On the other hand, the positive correlations were detected between *Candidatus Methanoplasma termitum* and *Prevotella, Porphyromonas, Bacillus, Sphingobacterium,* and *Moraxella*, as well as between *Candidatus Methanomethylophilus alvus* and *Fibrobacter, Eubacterium, and Mageeibacillus* in the classification provided by Kraken (Figure 3.2d).

3.4 Discussion

In this study, the comparison of taxonomic outcomes of two pipelines, Mothur (developed by Li et al., 2016) and Kraken (developed by Wood and Salzberg, 2014, and adapted to the conditions of this study), was performed to determine which is a better approach in rumen microbial classification when total RNA-seq data were used. The advent of high-throughput sequencing has greatly advanced our knowledge of the ecology and functional capacity of rumen microbes and their role in converting low-quality and unusable feedstuffs into energy sources for host productivity (McCann et al., 2017). As a result, an assiduous effort has been made to unveil the linkage between the rumen microbiota and phenotypic traits of interest such as feed efficiency (Li and Guan, 2017), enzyme discovery (Qi et al., 2011) and methane emissions (Kittelmann et al., 2014; Shi et al., 2014; Kamke et al., 2016). Metagenomic studies have shown that the host may regulate the microbiota and its metabolic activity in relation to feed efficiency (FCR) through host-microbiome cross talk genes such as *TSTA3* (*GDP-L-fucose synthetase*) and *Fucl* (*L-fucose isomerase*), suggesting that the relative abundance of these genes could be used as a predictor for host feed efficiency (Roehe et al., 2016). Although the number of rumen metagenomics and metatranscriptomics studies has grown enormously over the last couple of years (McCann et al., 2017), the functional outcomes and biological interpretation of omics data strongly depend on the computational methods used (Simon and Daniel, 2011; Siegwald et al., 2017). In this study, both Mothur and Kraken pipelines showed the rumen of the bulls to be dominated by *Prevotella*, *Treponema*, *Ruminoccocus*, *Fibrobacter*, and *Butyrivibrio*, which are considered as part of a “core bacterial microbiome” (Henderson et al., 2015). In addition to the mutual “core microbiome” shared by the two pipelines at the genus level, Kraken detected a relatively high abundance of 1) *Prevotella ruminicola* (Supplementary Table 2), which is involved in the ruminal digestion of hemicellulose and pectin (Marounek and Duskova, 1999); 2) *Fibrobacter succinogenes* (Supplementary Table 2), a gram-negative, fiber degrader species (Suen et al., 2011); and 3)

non-motile species within the *Ruminoccocus* genus (Supplementary Table 2) that share different niches (La Reau et al., 2016): *R. bicirculans,* which selectively utilizes hemicelluloses but not cellulose or arabinoxylan (Wegmann et al., 2014), and *R. albus*, which is capable of digesting cellulose and xylan (Christopherson et al., 2014).

Interestingly, both methods identified about 1% of Cyanobacteria (Supplementary Tables 1 and 2), corroborating the findings of previous studies that have reported low abundances of these oxygenic phototrophic bacteria in the rumen of dairy (Scharen et al., 2017) and beef cattle (Li and Guan, 2017), and of camels (Gharechahi et al., 2015). Cyanobacteria are aerobic bacteria that can perform carbohydrate fermentation in a deficient N_2 concentration (heterocystous) or in a combination of N_2 deficiency and anoxic conditions (nonheterocystous) (Nandi and Sengupta, 1998). Although the ruminal environment is widely considered to be anaerobic, significant concentrations of O_2 (60 and 100 nmol/min per mL) can be detected in the rumen fluid (Newbold et al., 1996), indicating that the presence of Cyanobacteria in the rumen may be related to O_2 scavenging and sugar fermentation performed under restrict aerobic conditions. It is important to mention that although Cyanobacteria has been widely detected in aqueous and soil environments (Williams et al., 2004; Cruz-Martinez et al., 2009), the identification of this phylum in the mammals' gut has raised critical questions on what roles these organisms may play in aphotic and anaerobic habitats (Soo et al., 2014) like the rumen. Recent researches have reported that gut Cyanobacteria are highly conserved but their 16S rRNA gene phylogenetic tree differed from the photosynthetic Cyanobacteria, which led to the designation of a new candidate class called Melainabacteria (Soo et al., 2014) whose members are capable of fermenting a range of sugars (e.g., glucose, fructose, sorbitol) into acetate and butyrate in the gut (Di Rienzi et al., 2013). Neither Kraken nor Mothur identified Melainabacteria in the samples, demonstrating that further studies are needed to disentangling its role in the rumen.

However, the two methods (Kraken and Mothur) generated microbial classification at different taxonomic levels for rumen bacteria. To completely understand the function of the rumen microbiota, it is essential to identify organisms at the species level since different species, within the same genus, can have varied functions and niches. The Mothur based method was useful to identify a diverse bacterial microbiota from the RNA-seq datasets, but it was not able to classify any of the bacterial sequences further than the genus level (Tables 3.1 and 3.2). Microbial classification to the species level is a major challenge for clustering-

first approaches based on targeted regional 16S rRNA when short (up to 250 bp) or even longer reads generated from total RNA-seq are used to identify environmental microbes (Xiang et al., 2017). Most existing tools (for bacteria and archaea) lack solid probabilistic-based criteria to evaluate the accuracy of taxonomic assignments to determine the best-matched database hits to distinguish multiple species from the targeted sequence region of the 16S rRNA gene (Xiang et al., 2017). To identify bacteria at the species level, sequencing of full length of 16S rRNA is desired and thus future studies need to increase the sequence length to enhance the resolution for microbial identification. For the Kraken based approach, the reference database was built based on all known microbial genomes and as a result it generated a higher resolution (to the species level) of the rumen microbiota, enabling the program to annotate each microbial sequence to the LCAs (Wood and Salzberg, 2014). In this process, k-mer paths formed by Kraken assign a specific weight to each node (equal to the number of sequences associated with the node's taxon) while increasing the sensitivity of the species classification even if regions (for example, V3-V5) of the 16S rRNA gene were analyzed (Wood and Salzberg, 2014; Valenzuela-González et al., 2016). Consequently, the generation of chimeric trees using short or long input sequences is improbable with Kraken as unlike other programs (such as Ribosomal Database Project classifier and Mothur), it leaves out specific sequences if there is insufficient evidence for classification and they are designated as unclassified (Valenzuela-González et al., 2016). Therefore, inputting short or long environmental sequences (containing most of the 16S hypervariable regions or mRNA sequences) into Kraken may generate a more representative profile of complex microbiomes (Valenzuela-González et al., 2016) like the rumen. However, the lack of reference genomes for rumen microorganisms also limits Kraken. For example, the classification of *Xenorhabdus* (Table 2) and *Xenorhabdus doucetiae* (data not shown; relative abundance (%): H-FCR, 0.1 ± 0.10 found in 6 samples; L-FCR, 0%), a motile, gram-negative soil bacterium usually described as being part of entomopathogenic nematode/bacterium symbiotic complex (Furgani et al., 2008) has not been previously reported in amplicon based sequencing (Li et al, 2016a) or metagenomic/metatranscriptome sequencing (Li and Guan, 2017) of rumen contents. The classification of this bacterial species may indicate that Kraken did not properly identify the microbe since the reference genome information was built mostly from all microbial genomes annotated in the NCBI database. However, these organisms may have been actually detected in the rumen since cattle can consume soil, raising the possibility that their detection was transitory.

It is noteworthy that Methanobrevibacter (family Methanobacteriaceae) was identified in both databases (Supplementary Tables 1, 2, and 4). This genus has been reported to be the most abundant archaeal population in the rumen based on DNA datasets (Kittelmann et al., 2013; Henderson et al., 2015), but it had a lower abundance than Methanomassiliicoccaceae at the RNA level in this study. This result is consistent with the research conducted by Li et al. (2016a), who reported a predominance of Methanomassiliicoccaceae over Methanobrevibacter in RNA-based datasets when compared to DNA Amplicon-seq outcomes, suggesting that Methanomassiliicoccaceae may be more active in the rumen than Methanobacteriaceae. However, further studies are needed to determine whether the differences in abundance between those two archaeal populations have a methodological influence or are controlled by diet, host animal or management strategies. Unlike bacterial classification, Kraken and Mothur generated contrasting results on archaea identification (Table 3.3), which reflects the divergent taxonomic profiles at the species level. For example, certain archaeal genomes, such as *Methanobrevibacter wolinii* and *Methanobrevibacter woesei*, were only found in the rumen-specific archaea database, as the Kraken standard database lacked these complete genomes. However, Kraken was able to detect *Candidatus Methanoplasma termitum* and *Candidatus Methanomethylophilus alvus*, which were not identified by Mothur pipeline. Li et al. (2016b) isolated the archaeon ISO4-H5 (member of the order Methanomassiliicoccales) from the sheep rumen and discovered that this archaeal taxon exhibited genome size (1.9 Mb) and GC content (54%) similar to *Candidatus Methanoplasma termitum* (enriched from the termite gut) and *Candidatus Methanomethylophilus alvus* (enriched from human feces). These two species encode pathways required for hydrogen-dependent methylotrophic methanogenesis by reduction of methyl substrates, without the ability to oxidize methyl substrates to carbon dioxide (Li et al., 2016b). Thus, it is possible that these microbes reside in the rumen. Future analysis with archaeon ISO4-H5 sequences included in the databases of both pipelines as well as its isolation, culture and characterization may provide further evidence of this possibility.

To further verify how these two methods affected data interpretation, the rumen microbiota of H-FCR and L-FCR bulls were compared based on the taxonomic outcomes generated by the two software packages. Both computational pipelines revealed differences in microbial abundance between H- and L-FCR groups at all taxonomic ranks, with Mothur

exclusively identifying a higher abundance of poorly characterized bacterial phylotypes (e.g., R4-41B) in L-FCR bulls (Table 3.2). It has been reported that the abundance of R4-41B was negatively correlated with production traits over the first 12 weeks postpartum in dairy cows (Lima et al., 2015), suggesting that it may have undesirable impacts on the function of the rumen microbiome of L-FCR cattle. Although Kraken identified a relatively higher abundance of *Xenorhabdus* in H-FCR bulls (Table 2), this result could be erroneous with further validation needed as described above. However, researchers have enumerated and identified a high number (15.7×10^4 Most Probable Number/g) of chlortetracycline resistant Enterobacteriacea in cattle feces that largely consisted of *Xenorhabdus doucetiae* (Watanabe et al., 2016). Since antimicrobial agents (e. g., chlortetracycline) are typically administered subtherapeutically to beef cattle (Inglis et al., 2005), our results suggest that H-FCR animals may be more susceptible to harbor chlortetracycline resistant bacteria than L-FCR animals in the event of a therapeutic administration of this antibiotic. Further investigations aiming to evaluate the effects of antimicrobial agents (e. g., chlortetracycline) on the development of antimicrobial resistance in *Xenorhabdus* recovered from less efficient cattle (H-FCR) are warranted. Kraken also detected a higher (P = 0.09) abundance of Methanococcaceae (relative abundance (%): H-FCR, 13.6 ± 8.96; L-FCR, 4.1 ± 4.85) in the rumen of H-FCR bulls, indicating that Methanococcaceae may play a potential role in the linkages between methanogenesis and reduced feed efficiency in cattle. Although RNA-targeted DNA probes and genomic DNA sequencing have revealed a significant population of this archaeal family residing in the rumen (Janssen and Kirs, 2008) and exhibiting a positive correlation with increased forage content in the diet (Pitta et al., 2016), members of this methanogenic archaea family still need to be cultured from the rumen to test our findings.

Finally, our study demonstrated that both pipelines (Mothur and Kraken) were effective in detecting a lower bacterial diversity in efficient (L-FCR) cattle (Table 3.4), corroborating the recent findings by Li and Guan et al. (2017) and (Shabat et al., 2016) that the rumen microbiota of efficient cattle is less complex and more specialized in harvesting energy from the diet through simpler metabolic networks (e.g., acrylate pathway) than inefficient cattle. However, only Kraken identified a significantly lower diversity in the archaeal community in L-FCR bulls, but this result should be carefully interpreted as many archaea phylotypes classified by Kraken are environmental organisms that have not yet been described in the rumen. For example, the methane-producing archaeon *Methanothermococcus okinawensis*

(the third-most abundant archaea taxon classified by Kraken, Supplementary Table 2) was first isolated from a deep-sea hydrothermal vent system (Takai et al., 2002), *Picrophilus torridus* and *Acidilobus saccharovorans* (the fourth and fifth-most abundant archaea taxa detected by Kraken, Supplementary Table 2) were isolated from a dry solfataric field (Fütterer et al., 2004) and a terrestrial acidic hot spring (Mardanov et al., 2010), respectively. Thus, it is worth mentioning that, in spite of the Kraken's promising results, this pipeline is severely limited when studying a microbiome that is not well described in its standard database (like the rumen), indicating that Mothur (using a specific archaea database described by Li et al., 2016a) could be more suited for identifying archaeal taxonomic profiles.

3.5 Conclusion

The current study was the first to compare the molecular-phylogenetic outcomes of Mothur and Kraken using transcriptomic sequence data (~140 bp in length) of rumen samples. The Kraken pipeline has been adapted to include reference genomes for rumen specific organisms, which has led to the identification of rumen bacteria at species level and more bacterial phylotypes. However, the results of the archaeal classification as well as some of the bacterial species identified by Kraken should be carefully interpreted as many detected phylotypes have not yet been described in the rumen, highlighting the importance of strengthening the Kraken database through the inclusion of more genomes annotated by single cell sequencing of rumen cultures/isolates to enable a more accurate classification. As to the future directions, new sequenced genomes (410 draft bacterial and archaeal genomes) by Hungate1000 project (JGI database) will be included in the Kraken standard database and the recently developed Rumen and Intestinal Methanogen Database will be used for archaea classification (Seedorf et al., 2014), with the goal of improving the accuracy of the results. It is also proposed the configuration of a joint pipeline using both Kraken and Mothur simultaneously to improve the resolution of taxonomic profiling of the rumen microbiome. This joint pipeline will produce a final rumen microbial profile obtained from the combination of multiple results generated from different bioinformatics tools as outlined by Piro et al. (2017), who published a computational method called *MetaMeta* that executes and integrates results from six metagenomic analysis tools (CLARK - Ounit et al., 2015; DUDes - Piro et al., 2016; GOTTCHA - Freitas et al., 2015; KRAKEN - Wood and Salzberg, 2014; KAIJU - Menzel et al., 2016; and mOTUs - Sunagawa et al., 2013). If the rumen microbiome datasets are strengthened to the same level as the human databases, the joint pipeline will generate more

sensitive and reliable results than those of the best single profile (generated separately by each tool) (Piro et al. 2017). It is believed that a joint pipeline supported by a collection of tools will be useful to control sources of variation present in any metagenomics/metatranscriptomic analysis (e.g., analytical pipelines, related databases and software parameters), which will ultimately lead to standardized results and more reliable biological interpretations. In addition, although Kraken has improved the taxonomic assessment at species level, the high number of unclassified sequences (65%) suggests a need for identifying the rumen microbes with a more resolved taxonomy assignment. Regardless of the approach undertaken, the only way for improvement is through a continued strengthening of the databases by including additional information of whole genome sequencing of rumen isolates as well as single cell sequencing of unculturable rumen microbes, as the ability to culture rumen microorganisms is still limited.

3.6 Literature cited

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society Series B-Methodological* 44(2)**,** 139-177.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* 57(1)**,** 289-300.

Béra-Maillet, C., Mosoni, P., Kwasiborski, A., Suau, F., Ribot, Y., and Forano, E. (2009). Development of a RT-qPCR method for the quantification of Fibrobacter succinogenes S85 glycoside hydrolase transcripts in the rumen content of gnotobiotic and conventional sheep. *Journal Of Microbiological Methods* 77(1)**,** 8-16. doi: 10.1016/j.mimet.2008.11.009.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15)**,** 2114-2120. doi: 10.1093/bioinformatics/btu170.

Cao, K.A.L., Costello, M.E., Lakis, V.A., Bartolo, F., Chua, X.Y., Brazeilles, R., et al. (2016). MixMC: A Multivariate Statistical Framework to Gain Insight into Microbial Communities. *Plos One* 11(8). doi: 10.1371/journal.pone.0160169.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7(5)**,** 335-336. doi: 10.1038/nmeth.f.303.

Chen, J., and Li, H.Z. (2013). Variable selection for sparse dirichlet-multinomial regression with an application to microbiome data analysis. *Annals of Applied Statistics* 7(1)**,** 418-442. doi: 10.1214/12-aoas592.

Chomczynski, P., and Sacchi, N. (2006). The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nature Protocols* (2)**,** 581.

Christopherson, M.R., Dawson, J.A., Stevenson, D.M., Cunningham, A.C., Bramhacharya, S., Weimer, P.J., et al. (2014). Unique aspects of fiber degradation by the ruminal

ethanologen Ruminococcus albus 7 revealed by physiological and transcriptomic analysis. *Bmc Genomics* 15. doi: 10.1186/1471-2164-15-1066.

Cruz-Martinez, K., Suttle, K.B., Brodie, E.L., Power, M.E., Andersen, G.L., and Banfield, J.F. (2009). Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *Isme Journal* 3(6)**,** 738-744. doi: 10.1038/ismej.2009.16.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72(7)**,** 5069-5072. doi: 10.1128/aem.03006-05.

Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., et al. (2013). The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *Elife* 2. doi: 10.7554/eLife.01102.

Duffield, T., Plaizier, J.C., Fairfield, A., Bagg, R., Vessie, G., Dick, P., et al. (2004). Comparison of techniques for measurement of rumen pH in lactating dairy cows. *Journal of Dairy Science* 87(1)**,** 59-66.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16)**,** 2194-2200. doi: 10.1093/bioinformatics/btr381.

Freitas, T.A.K., Li, P.-E., Scholz, M.B., and Chain, P.S.G. (2015). Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. *Nucleic Acids Research* 43(10)**,** e69-e69. doi: 10.1093/nar/gkv180.

Furgani, G., Böszörményi, E., Fodor, A., Máthé-Fodor, A., Forst, S., Hogan, J.S., et al. (2008). Xenorhabdus antibiotics: a comparative analysis and potential utility for controlling mastitis caused by bacteria. *Journal of Applied Microbiology* 104(3)**,** 745-758. doi: 10.1111/j.1365-2672.2007.03613.x.

Fütterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., et al. (2004). Genome sequence of Picrophilus torridus and its implications for life around pH 0. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 101(24)**,** 9091-9096.

Gharechahi, J., Zahiri, H.S., Noghabi, K.A., and Salekdeh, G.H. (2015). In-depth diversity analysis of the bacterial community resident in the camel rumen. *Systematic and Applied Microbiology* 38(1)**,** 67-76. doi: 10.1016/j.syapm.2014.09.004.

Gonzalez, I., Dejean, S., Martin, P.G.P., and Baccini, A. (2008). CCA: An R package to extend canonical correlation analysis. *Journal of Statistical Software* 23(12)**,** 1-14.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P.H., et al. (2016). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports* 6. doi: 10.1038/srep19175.

Inglis, G.D., McAllister, T.A., Busz, H.W., Yanke, L.J., Morck, D.W., Olson, M.E., et al. (2005). Effects of subtherapeutic administration of antimicrobial agents to beef cattle on the prevalence of antimicrobial resistance in Campylobacter jejuni and Campylobacter hyointestinalis. *Applied and Environmental Microbiology* 71(7)**,** 3872-3881. doi: 10.1128/aem.71.7.3872-3881.2005.

Janssen, P.H., and Kirs, M. (2008). Structure of the archaeal community of the rumen. *Applied & Environmental Microbiology* 74(12)**,** 3619-3625. doi: 10.1128/AEM.02812-07.

Jost L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology* 88, 2427–2439. 10.1890/06-1736.1

Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., et al. (2016). Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4(1)**,** 56-56.

Kittelmann, S., Seedorf, H., Walters, W.A., Clemente, J.C., Knight, R., Gordon, J.I., et al. (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *Plos One* 8(2). doi: 10.1371/journal.pone.0047879.

Kittelmann, S., Pinares-Patino, C.S., Seedorf, H., Kirk, M.R., Ganesh, S., McEwan, J.C., et al. (2014). Two different bacterial community types are linked with the low-methane emission trait in sheep. *PLoS ONE* 9(7)**,** 1-9. doi: 10.1371/journal.pone.0103171.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics (Oxford, England)* 28(24)**,** 3211-3217. doi: 10.1093/bioinformatics/bts611.

Korpela K., Salonen A., Virta L. J., Kekkonen R. A., Forslund K., Bork P., et al. (2016). Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nature Communication* 7:10410. 10.1038/ncomms10410.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied And Environmental Microbiology* 79(17)**,** 5112-5120. doi: 10.1128/AEM.01043-13.

La Reau, A.J., Meier-Kolthoff, J.P., and Suen, G. (2016). Sequence-based analysis of the genus Ruminococcus resolves its phylogeny and reveals strong host association. *Microbial Genomics* 2(12)**,** e000099-e000099. doi: 10.1099/mgen.0.000099.

Li, D.H., Liu, C.M., Luo, R.B., Sadakane, K., and Lam, T.W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31(10)**,** 1674-1676. doi: 10.1093/bioinformatics/btv033.

Li, F., and Guan, L.L. (2017). Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Applied And Environmental Microbiology*. doi: 10.1128/AEM.00061-17.

Li, F., Henderson, G., Sun, X., Cox, F., Janssen, P.H., and Guan, L.L. (2016a). Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches. *Frontiers in Microbiology* 7(987). doi: 10.3389/fmicb.2016.00987.

Li, Y., Leahy, S.C., Jeyanathan, J., Henderson, G., Cox, F., Altermann, E., et al. (2016b). The complete genome sequence of the methanogenic archaeon ISO4-H5 provides insights into the methylotrophic lifestyle of a ruminal representative of the Methanomassiliicoccales. *Standards in Genomic Sciences* 11. doi: 10.1186/s40793-016-0183-5.

Lima, F.S., Oikonomou, G., Lima, S.F., Bicalho, M.L.S., Ganda, E.K., Filho, J.C.d.O., et al. (2015). Prepartum and postpartum rumen fluid microbiomes: characterization and correlation with production traits in dairy cows. *Applied And Environmental Microbiology* 81(4)**,** 1327-1337.

Lindgreen, S., Adair, K.L., and Gardner, P.P. (2016). An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports* 6**,** 19233-19233. doi: 10.1038/srep19233.

Mandal, S., Van Treuren, W., White, R.A., Eggesbo, M., Knight, R., and Peddada, S.D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease* 26**,** 27663-Article No.: 27663. doi: 10.3402/mehd.v26.27663.

Mardanov, A.V., Svetlitchnyi, V.A., Beletsky, A.V., Prokofeva, M.I., Bonch-Osmolovskaya, E.A., Ravin, N.V., et al. (2010). The genome sequence of the crenarchaeon Acidilobus saccharovorans supports a new order, Acidilobales, and suggests an important ecological role in terrestrial acidic hot springs. *Applied And Environmental Microbiology* 76(16)**,** 5652-5657. doi: 10.1128/AEM.00599-10.

McCann, J.C., Elolimy, A.A., and Loor, J.J. (2017). Rumen Microbiome, Probiotics, and Fermentation Additives. *The Veterinary Clinics Of North America: Food Animal Practice*. doi: 10.1016/j.cvfa.2017.06.009.

Menzel, P., Ng, K.L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7**,** 11257-11257. doi: 10.1038/ncomms11257.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., et al. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9. doi: 10.1186/1471-2105-9-386.

Montanholi, Y.R., Swanson, K.C., Palme, R., Schenkel, F.S., McBride, B.W., Lu, D., et al. (2010). Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. *Animal* 4(5)**,** 692-701. doi: 10.1017/s1751731109991522.

Marounek, M., and Duskova, D. (1999). Metabolism of pectin in rumen bacteria Butyrivibrio fibrisolvens and Prevotella ruminicola. *Letters in applied microbiology* 29(6)**,** 429-433.

Nandi, R., and Sengupta, S. (1998). Microbial production of hydrogen: An overview. *Critical Reviews in Microbiology* 24(1)**,** 61-84. doi: 10.1080/10408419891294181.

Newbold, C.J., Wallace, R.J., and McIntosh, F.M. (1996). Mode of action of the yeast Saccharomyces cerevisiae as a feed additive for ruminants. *British Journal of Nutrition* 76(2)**,** 249-261. doi: 10.1079/bjn19960029.

Olfert E. D., Cross B. M., McWilliams A. A. (1993). Guide to the Care and Use of Experimental Steers. Ottawa, ON: Canadian Council on Animal Care.

Ounit, R., Wanamaker, S., Close, T.J., and Lonardi, S. (2015). CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics* 16(1)**,** 1-13. doi: 10.1186/s12864-015-1419-2.

Piro, V.C., Lindner, M.S., and Renard, B.Y. (2016). DUDes: a top-down taxonomic profiler for metagenomics. *Bioinformatics (Oxford, England)* 32(15)**,** 2272-2280. doi: 10.1093/bioinformatics/btw150.

Piro, V.C., Matschkowski, M., and Renard, B.Y. (2017). MetaMeta: integrating metagenome analysis tools to improve taxonomic profiling. *Microbiome* 5(1)**,** 101-101. doi: 10.1186/s40168-017-0318-y.

Pitta, D.W., Indugu, N., Kumar, S., Vecchiarelli, B., Sinha, R., Baker, L.D., et al. (2016). Metagenomic assessment of the functional potential of the rumen microbiome in Holstein dairy cows. *Anaerobe* 38**,** 50-60. doi: 10.1016/j.anaerobe.2015.12.003.

Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., et al. (2011). Snapshot of the eukaryotic gene expression in muskoxen rumen-a metatranscriptomic approach. *Plos One* 6(5). doi: 10.1371/journal.pone.0020521.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41(Database issue)**,** D590-D596. doi: 10.1093/nar/gks1219.

R Core Team (2016). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Roehe, R., Dewhurst, R.J., Duthie, C. A., Rooke, J.A., McKain, N., Ross, D.W., et al. (2016). Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. *Plos Genetics* 12(2). doi: 10.1371/journal.pgen.1005846.

Russell, J.B., and Rychlik, J.L. (2001). Factors that alter rumen microbial ecology. *Science* 292(5519)**,** 1119-1122. doi: 10.1126/science.1058830.

Scharen, M., Drong, C., Kiri, K., Riede, S., Gardener, M., Meyer, U., et al. (2017). Differential effects of monensin and a blend of essential oils on rumen microbiota composition of transition dairy cows. *Journal of Dairy Science* 100(4)**,** 2765-2783. doi: 10.3168/jds.2016-11994.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied & Environmental Microbiology* 75(23)**,** 7537-7541. doi: 10.1128/AEM.01541-09.

Seedorf, H., Kittelmann, S., Henderson, G., and Janssen, P.H. (2014). RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. *Peerj* 2. doi: 10.7717/peerj.494.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* 9(8)**,** 811-814. doi: 10.1038/nmeth.2066.

Shabat, S.K.B., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., Berg Miller, M.E., et al. (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J*. doi: 10.1038/ismej.2016.62.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11)**,** 2498-2504. doi: 10.1101/gr.1239303.

Shi, W., Kang, D., Froula, J., Fan, C., Deutsch, S., Chen, F., et al. (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Research* 24(9)**,** 1517-1525. doi: 10.1101/gr.168245.113.

Siegwald, L., Touzet, H., Lemoine, Y., Hot, D., Audebert, C., and Caboche, S. (2017). Assessment of Common and Emerging Bioinformatics pipelines for targeted metagenomics. *PLoS One* 12(1)**,** 1-26. doi: 10.1371/journal.pone.0169563.

Soo, R.M., Skennerton, C.T., Sekiguchi, Y., Imelfort, M., Paech, S.J., Dennis, P.G., et al. (2014). An Expanded Genomic Representation of the Phylum Cyanobacteria. *Genome Biology and Evolution* 6(5)**,** 1031-1045. doi: 10.1093/gbe/evu073.

Simon, C., and Daniel, R. (2011). Metagenomic Analyses: Past and Future Trends. *Applied and environmental microbiology AEM* 77(4)**,** 1153-1161.

Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J., et al. (2011). The complete genome sequence of fibrobacter succinogenes s85 reveals a cellulolytic and metabolic specialist. *PLoS One* 6. doi: e0018814.

Sunagawa, S., Mende, D.R., Zeller, G., Izquierdo-Carrasco, F., Berger, S.A., Kultima, J.R., et al. (2013). Metagenomic species profiling using universal phylogenetic marker genes. *Nature Methods* 10(12)**,** 1196-1199. doi: 10.1038/nmeth.2693.

Takai, K., Inoue, A., and Horikoshi, K. (2002). Methanothermococcus okinawensis sp. nov., a thermophilic, methane-producing archaeon isolated from a Western Pacific deep-sea hydrothermal vent system. *International Journal Of Systematic And Evolutionary Microbiology* 52(Pt 4)**,** 1089-1095.

Thompson, S. (2015). *The effect of diet type on residual feed intake and the use of infrared thermography as a method to predict efficiency in beef bulls.* Master's thesis. Winnipeg, MB: University of Manitoba.

Valenzuela-González, F., Martínez-Porchas, M., Villalpando-Canchola, E., and Vargas-Albores, F. (2016). Studying long 16S rDNA sequences with ultrafast-metagenomic sequence classification using exact alignments (Kraken). *Journal of Microbiological Methods* 122**,** 38-42. doi: 10.1016/j.mimet.2016.01.011.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73(16)**,** 5261-5267. doi: 10.1128/aem.00062-07.

Watanabe, K., Horinishi, N., Matsumoto, K., Tanaka, A., and Yakushido, K. (2016). A New Evaluation Method for Antibiotic-Resistant Bacterial Groups in Environment. *Advances in Microbiology* 6(3)**,** 133-151. doi: 10.4236/aim.2016.63014.

Wegmann, U., Louis, P., Goesmann, A., Henrissat, B., Duncan, S.H., and Flint, H.J. (2014). Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota ('Ruminococcus bicirculans') reveals two chromosomes and a selective capacity to utilize plant glucans. *Environmental Microbiology* 16(9)**,** 2879-2890. doi: 10.1111/1462-2920.12217.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5. doi: 10.1186/s40168-017-0237.

Williams, M.M., Domingo, J.W.S., Meckes, M.C., Kelty, C.A., and Rochon, H.S. (2004). Phylogenetic diversity of drinking water bacteria in a distribution system simulator. *Journal of Applied Microbiology* 96(5)**,** 954-964. doi: 10.1111/j.1365-2672.2004.02229.x.

Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15(3)**,** R46-R46. doi: 10.1186/gb-2014-15-3-r46.

Xiang, G., Huaiying, L., Kashi, R., and Qunfeng, D. (2017). A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics* 18**,** 1-10. doi: 10.1186/s12859-017-1670-4.

3.7 Tables and figures

Table 3.1 Quantification of taxonomic phylotypes identified by each method

Phylotypes	Mothur[1]		Kraken[2]	Commonly Detected Phylotypes (N°)
	Classified (N°)	Unclassified (N°)	Classified (N°)	
Bacteria				
Phyla	23	1	26	16
Families	121	66	204	78
Genera	189	135	348	69
Species	-	-	423	0
Archaea				
Phyla	1	1	2	1
Families	3	3	7	2
Genera	4	5	8	1
Species	5	6	8	1

[1]Pipeline to assess the rumen microbiota developed by Li et al. (2016) based on Mothur (Schloss et al., 2009). Clustering-first approaches (such as Mothur) allow the discrimination of unclassified reads (Siegwald et al., 2017).

[2]Metagenomic sequence classification method developed by Wood and Salzberg (2014). Unlike clustering-first approaches, assignment-first tools (such as Kraken) do not allow the discrimination of unclassified reads (Siegwald et al., 2017).

Table 3.2 Differentially abundant bacteria in efficient (low FCR) and inefficient (high FCR) cattle according to the two classification methods[1,2,3]

Phylotypes	Mothur		Kraken	
	High (%)	Low (%)	High (%)	Low (%)
Phyla				
Bacteroidetes	35.0 ± 8.52	43.2 ± 6.98	37.7 ± 9.75	46.1 ± 11.35
Firmicutes	24.8 ± 9.53	19.2 ± 6.71	28.1 ± 7.01	22.6 ± 6.98
Proteobacteria	20.0 ± 4.88	14.6 ± 3.12	15.7 ± 2.59	12.8 ± 2.63
Fibrobacteres	2.50 ± 0.80	4.4 ± 1.28	1.1± 0.42	1.3 ± 0.31
Spirochaetes	2.0 ± 0.50	2.5 ± 0.33	4.6 ± 0.60	5.1 ± 1.02
Verrucomicrobia	1.4 ± 0.19	2.1 ± 0.82	2.1 ± 0.42	2.4 ± 0.62
Families				
Prevotellaceae	18.4 ± 6.73	23.7 ± 6.17	26.9 ± 10.48	35.3 ± 13.63
Ruminococcacea	10.2 ± 4.38	7.73 ± 3.89	8.4 ± 3.01	5.7 ± 3.33
Lachnopiraceae	7.1 ± 3.62	5.5 ± 2.23	7.1 ± 1.65	6.1 ± 1.95
Fibrobacteriacea	2.6 ± 0.83	4.6 ± 1.34	1.8 ± 0.37	1.67 ±0.51
Spirochaetaceae	1.8 ± 0.56	2.3 ± 0.34	4.8 ±0.73	5.3 ±1.21
R4 – 41B	0.02 ± 0.037[a]	0.13 ± 0.118[b]	-	-
Actinomycetaceae	-	-	0.1± 0.03 [b]	0.2 ±0.04 [a]
Genera				
Prevotella	20.0 ± 8.28	25.2 ± 7.10	28.8 ± 10.75	37.5 ± 14.15
Ruminococcus	5.9 ± 2.88	4.4 ± 2.56	3.9 ± 1.70	2.7 ± 1.83
Fibrobacter	3.2 ± 1.12	5.5 ± 1.56	1.4 ± 0.56	1.6 ± 0.34
Butyrivibrio	1.0 ± 0.19	1.3 ± 0.68	2.2 ± 0.29	2.5 ± 1.31
Xenorhabdus	-	-	0.29 ± 0.246[a]	0.04 ± 0.036[b]

Species				
Prevotella ruminicola	-	-	23.0 ± 9.99	31.6 ± 13.15
Butyrivibrio proteoclasticus	-	-	2.6 ± 0.33	2.9 ± 1.44
Ruminiclostridium sp KB18	-	-	2.8 ± 0.98	1.6 ± 1.14
Fibrobacter succinogens	-	-	1.6 ± 0.66	1.8 ± 0.37
Ruminococcus albus	-	-	1.7 ± 0.89	1.0 ± 0.77

[1]Statistical comparisons were obtained by the application of ANCOM (Mandal et al., 2015) on taxa counts determined by Mothur (OUTs) or Kraken (K-mers), and thus estimators are comparable only between High (n=6) and Low (n=6) FCR cattle as provided by each classification method.

[2]P values were obtained using Wilcoxon exact test (calculated on the log-ratio matrix; Mandal et al., 2015), and then adjusted to FDR using Benjamini-Hochberg algorithm (Benjamini and Hochberg, 1995). A threshold of FDR < 0.15 was applied to determine significance. Within a row, means with different superscript are statistically different between High and Low FCR cattle for each method (separately).

[3]Blank spaces indicate that the phylotypes were not detected in the dataset either by Mothur or Kraken.

Table 3.3 Differentially abundant archaea in efficient (low FCR) and inefficient (high FCR) cattle according to the two classification methods[1,2,3]

Phylotypes	Mothur		Kraken	
	High (%)	Low (%)	High (%)	Low (%)
Families				
RCC and relatives	73.2 ± 3.77	71.9 ± 13.12	-	-
Methanomassiliicoccaceae	-	-	65.5 ± 9.92	67.1 ± 11.29
Methanococcaceae	-	-	13.6 ± 8.96[a]	4.1 ± 4.85[b]
Methanobacteriaceae	23.9 ± 4.19	24.8 ± 13.75	6.0 ± 6.42	7.0 ± 7.52
Methanosarcinaceae	0.3 ± 0.35	0.6 ± 0.72	5.6 ± 8.51	10.7 ± 5.32
Genera				
Candidatus Methanoplasma	-	-	49.0 ± 13.61	55.4 ± 9.51
Candidatus Methanomethylophilus	-	-	19.0 ± 7.92	12.8 ± 6.78
Methanosarcina	-	-	5.1 ± 9.20	11.0 ± 5.81
Methanobrevibacter	21.8 ± 3.85	21.8 ± 10.53	4.7 ± 5.01	5.3 ± 5.35
Methanosphaera	0.8 ± 0.47	1.2 ± 1.57	-	-
Methanimicrococcus	0.3 ± 0.33	0.6 ± 0.70	-	-
Species				
Candidatus Methanoplasma termitum	-	-	51.0 ± 14.27	62.8 ± 11.23
Candidatus Methanomethylophilus alvus	-	-	19.7 ± 8.28	14.6 ± 8.09
Methanobrevibacter gottschalkii and relatives	14.8 ± 3.60	14.7 ± 6.11	-	-
Methanobrevibacter ruminantium	3.8 ± 1.70	4.0 ± 4.19	2.4 ± 4.05	1.2 ± 3.14

Methanobrevibacter wolinii and relatives	0.1 ± 0.14	0.2 ± 0.32	-	-
Methanobrevibacter woesei	0.1 ± 0.10	0.05 ± 0.06	-	-
Methanobrevibacter smithii	0.03 ± 0.07	0.10 ± 0.14	-	-

[1]Statistical comparisons were obtained by the application of ANCOM (Mandal et al., 2015) on taxa counts determined by Mothur (OUTs) or Kraken (K-mer), and thus estimators are comparable only between High (n=6) and Low (n=6) FCR cattle as provided by each classification method.

[2]P values were obtained using Wilcoxon exact test (calculated on the log-ratio matrix; Mandal et al., 2015), and then adjusted to FDR using Benjamini-Hochberg algorithm (Benjamini and Hochberg, 1995). A threshold of FDR < 0.15 was applied to determine significance. Within a row, means with different superscript are statistically different only between High and Low FCR cattle for each method (separately).

[3]Blank spaces indicate that the phylotypes were not detected in the dataset either by Mothur or Kraken.

Table 3.4 Comparison of bacterial and archaeal alpha-diversity indexes between efficient (low FCR) and inefficient (high FCR) cattle according to the two microbial classification methods[1]

Indexes	Bacteria				Archaea			
	Mothur		Kraken		Mothur		Kraken	
	High	Low	High	Low	High	Low	High	Low
Number of observed phylotypes	244.1 ± 23.88	239.3 ± 19.98	241.1 ± 28.29	224.5 ± 28.37	8.8 ± 1.33	8.8 ± 0.75	5.0 ± 0.89	4.5 ± 1.05
Shannon[2]	2.78 ± 0.12^{a}	2.73 ± 0.14^{b}	3.93 ± 0.42^{a}	3.51 ± 0.62^{b}	0.90 ± 0.08	0.91 ± 0.28	1.27 ± 0.19^{a}	1.06 ± 0.25^{b}
Inverse Simpson	9.8 ± 1.93^{a}	8.7±1.82^{b}	12.7 ± 5.75^{a}	8.8 ± 5.60^{b}	1.74 ± 0.13	1.85 ± 0.61	2.95 ± 0.89^{a}	2.30 ± 0.64^{b}
Simpson (with rarefy)	0.89 ± 0.03^{a}	0.88 ± 0.03^{b}	0.90 ± 0.07^{a}	0.83 ± 0.11^{b}	0.42 ± 0.05	0.42 ± 0.15	0.64 ± 0.04^{a}	0.54 ± 0.05^{b}

[1]Within a row, means with different superscript were different at $P < 0.05$. Comparison was conducted using paired Wilcoxon signed rank test for bacteria (genus level) and archaea (species level) separately for High and Low FCR cattle as provided by each classification method, and thus estimators between bacterial and archaeal groups are comparable only within each method and between High and Low FCR animals.

[2]Shannon indices showed in the table are the raw values, and the comparison of Shannon indices between High and Low FCR cattle was based on the exponentially transformed values (Jost, 2007) using paired Wilcoxon signed rank test.

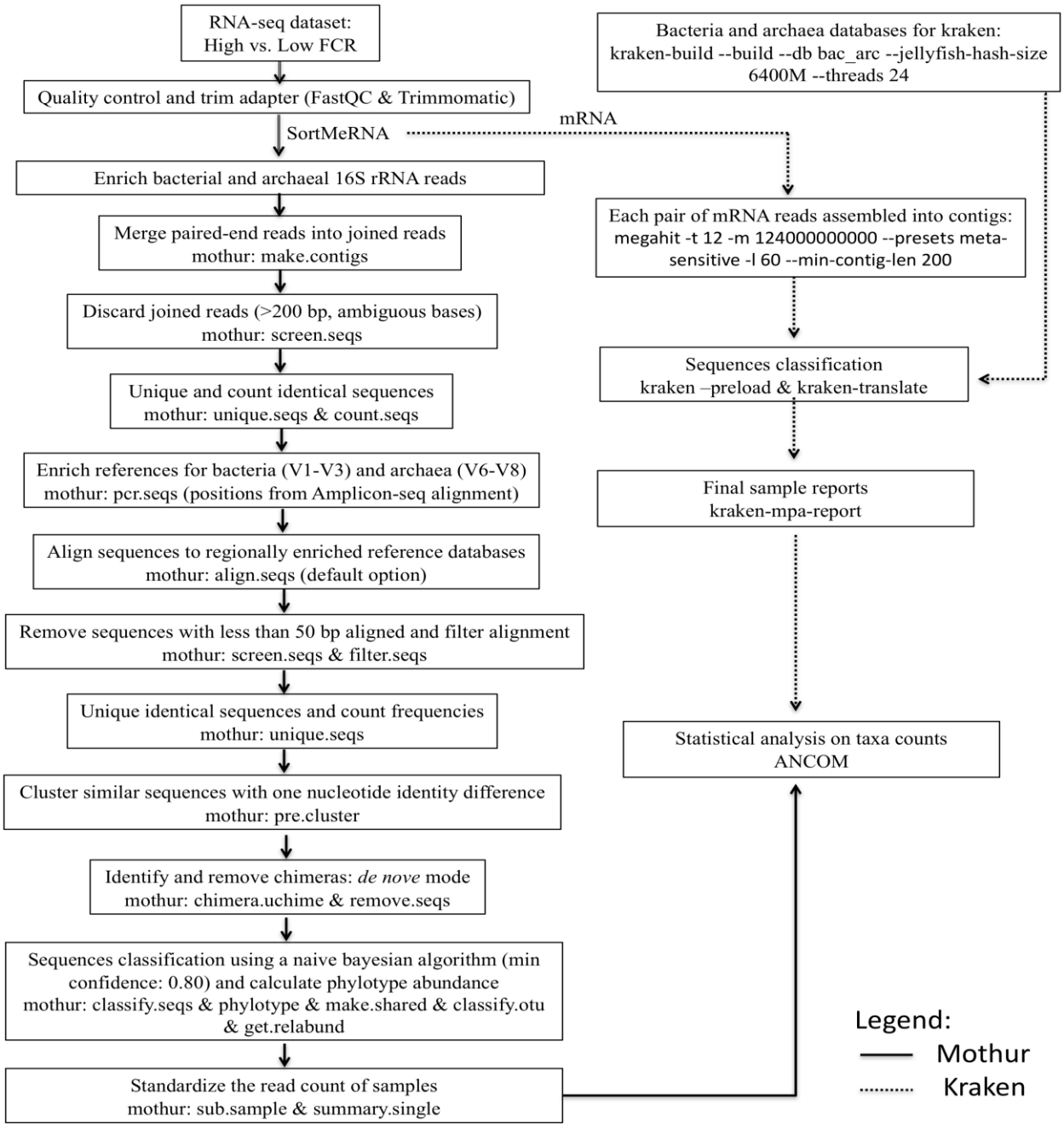


Figure 3.1 Flow chart of the pipelines (Mothur and Kraken) presenting software parameters used to analyze the rumen microbiota. Part of this figure was adapted from the pipeline published by Li et al. (2016a).

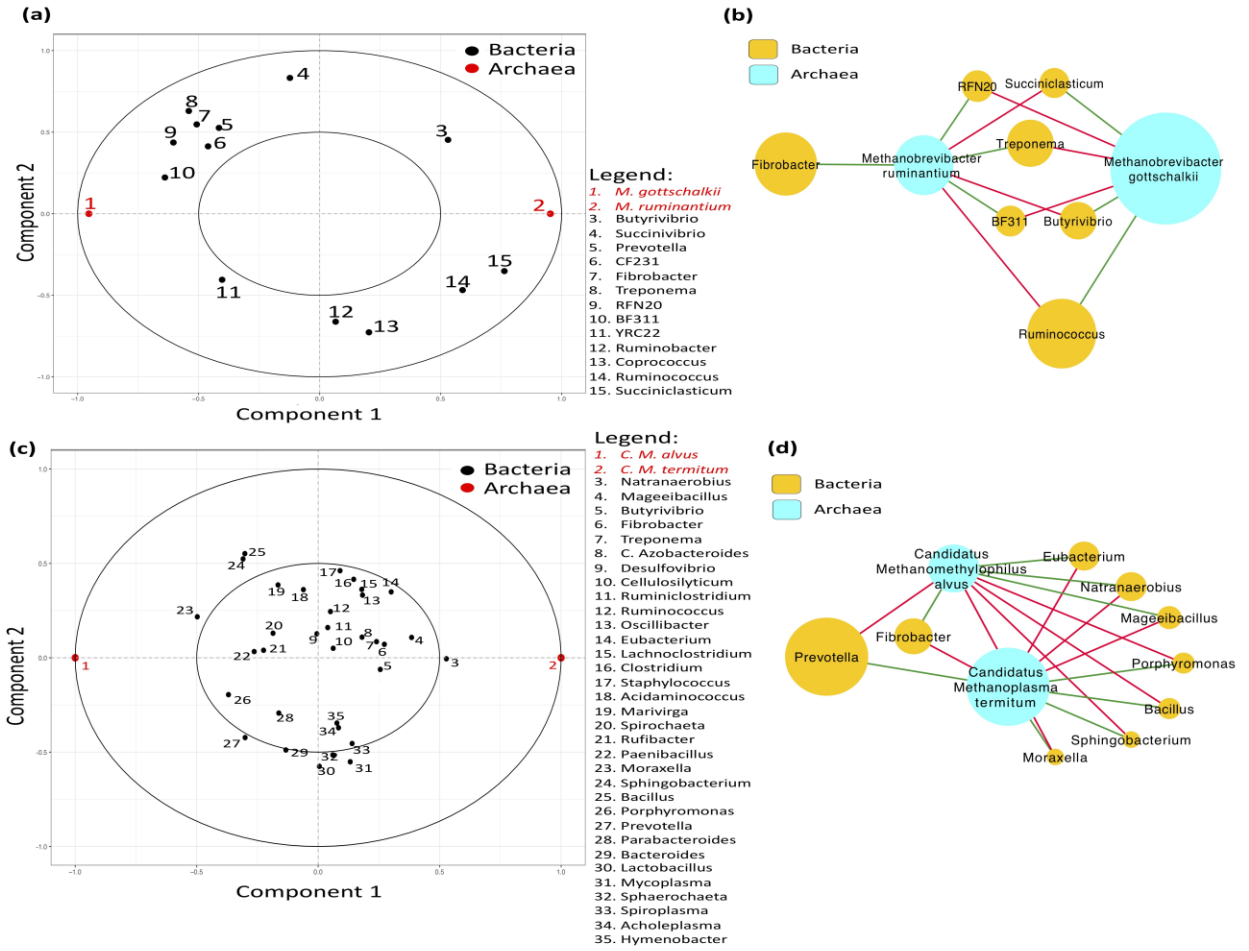


Figure 3.2 Correlation circle plots and relevance networks generated from the output of regularized canonical correlation (rCC) method (Total Sum Scaling + Centered Log Ratio) applied to rumen bacteria (genera) and archaea (species) classified by Mothur or Kraken. (a) and (b) show the correlation and network plots of the first two rCC components for Mothur. (c) and (d) represent the correlation and network plots of the first two rCC components for Kraken. In the correlation circle plots, bacteria (X) and archaea (Y) are shown inside a circle of radius 1 centered at the origin, with strongly associated (or correlated) variables being projected in the same direction from the origin. The greater the distance from the origin indicates stronger association. Two circumferences of radius 1 and 0.5 are plotted to reveal the correlation structure of the variables (Gonzalez et al., 2008). In the relevance networks, red and green edges indicate positive and negative

correlations respectively, and the sizes of the nodes indicate the mean average abundance. Only bacterial genera and archaeal species with a relative abundance > 0.1% detected in all rumen samples were included in the rCC analysis (Li and Guan, 2017).

Chapter 4

Taxonomic and functional assessment reveals the effect of Angus breed genetics on rumen microbial signatures

4.1 Introduction

The rumen is a complex anaerobic ecosystem inhabited by a symbiotic array of bacterial, archaeal, protozoal, and fungal species that work together to supply protein, vitamins and short-chain organic acids to the ruminant host (Russell and Rychlik 2001). Compared to other groups of microbes, bacteria represent the most genetically diverse and abundant organisms of the rumen microbiome (density of 10^{10}-10^{11} cells/ml rumen fluid), making up more than 50% of the cell mass (Opdahl et al. 2018, Creevey et al. 2014). Many bacteria species (e.g., *Prevotella ruminicola*, *Ruminococcus albus*, and *Fibrobacter succinogenes*) have been isolated and cultured from the rumen microbial community (Russell and Rychlik 2001), and efforts have been made to reveal functional activities of rumen bacteria with the purpose of enhancing feed efficiency of the host (Fuyong Li and Le Luo Guan 2017). The most important functional activity performed by rumen bacteria involves the digestion and metabolism of plant structural carbohydrates to aid in the breakdown of lignocellulose, as the ruminant host does not produce enzymes involved in plant cell wall digestion (Wang and McAllister 2002). Additionally, rumen bacteria have been associated with rumen fermentation parameters (e.g., volatile fatty acids - VFAs, NH_3-N, pH) and host performance (dry matter intake, average daily gain, and feed conversion ratio - FCR) (Zhou et al. 2018), suggesting that the identification of host-specific bacteria can contribute to improve our knowledge of microbial signatures associated with feed efficiency, and this may offer opportunities to enhance the efficiency of digestion in the rumen.

Long-standing efforts to increase the efficient use of feedstuffs to enhance ruminant growth have moved producers to adopt feed efficiency indicators like feed conversion ratio (FCR, feed/gain) to monitor phenotypic and genetic variations in host productivity (Crews 2005). More recently, a study showed that the host may regulate the rumen microbiota and its function in relation to FCR through microbial genes such

as *TSTA3* (*GDP-l-fucose synthetase*) and *FucI* (*l-fucose isomerase*), demonstrating that the relative abundance of these genes could serve as a predictor for host feed efficiency (Roehe et al. 2016). Roehe et al. (2016) also highlighted the effect of genetically distant crossbred breeds (Aberdeen Angus and Limousin) on the rumen microbiota and its microbial genes, suggesting that the host genetics plays a significant role in shaping the composition of the rumen microbiota. Studies have revealed that twins exhibited a different fecal microbial signature despite their genetic similarity (Lee et al. 2011) and variations in animal gain to feed ratios were observed in cattle fed and managed under the same environment although they had a similar genetic background (Wolfger et al. 2016). Based on these assumptions, this chapter speculated that variations in the rumen microbiota associated with the genetic makeup of the host could exist in genetically similar breeds.

Here, dimensionality reduction techniques (Cao et al. 2016) and metatranscriptomics were used to identify the subtle differences in the active rumen bacterial and functional signatures that characterize the closely related breeds of Black and Red Angus cattle. The genomic differences between these two beef breeds are minimal and attributable to genes (e.g., *MC1-R – melanocortin 1 receptor*) encoding proteins involved in biological functions of pigmentation (coat color) (Wolfger et al. 2016, McLean and Schmutz 2009). However, Black and Red Angus in terms of genetic diversity and population structure are described as separate breeds in North America (Márquez et al. 2010). Thus, comparisons between them with respect to the taxonomic and functional profiles of their bacterial population can ultimately help uncover microbial signatures that differentiate phenotypes of the Angus breed. Therefore, the objectives of this study were to (a) characterize active bacterial and microbial functional signatures discriminating two breeds of beef cattle (Black vs. Red Angus) fed forage-based diets, and (b) identify specific bacterial groups and functions associated with feed efficiency.

4.2 Materials and methods

4.2.1 Animal trial and sampling

The experimental procedures described here were reviewed and approved by the University of Manitoba animal care committee. Briefly, rumen contents were collected from 6 (Black Angus= 3; Red Angus= 3) purebred bulls (mean age of 249 ± 22 days and average body weight of 313.9 ± 32 kg) raised in confinement at the Glenlea Research Station (University of Manitoba) according to the guidelines of the Canadian Council on Animal Care. In the current trial, which lasted 180 days, bulls were fed a forage based diet (alfalfa hay, 17.9%; corn silage, 81.7%; limestone, 0.2%; salt, 0.1%; mineral, 0.1%) throughout the experimental period. Then, representative samples (250 ml) of rumen contents were collected over four-time points (0, 80, 100, 180 d) (24 samples in total) using a Geishauser oral probe (Geishauser 1993), immediately snap frozen in liquid nitrogen, and stored at -80°C for three months before RNA extraction. The feed intake of individual bulls was recorded using the GrowSafe® feeding system (GrowSafe Systems Ltd., Airdrie, Alberta, CA), and the FCR was calculated biweekly as a ratio of dry matter intake to average daily gain.

4.2.2 VFA analysis

Supernatants from rumen fluid samples were obtained after centrifugation at 3,000 × g for 15 min at 4˚C and mixed with 25% phosphoric acid (4:1; v/v) for the subsequent gas chromatography (GC) analysis. After adding the internal standard to the samples and incubating them at -20˚C overnight, they were centrifuged at 19,000 × g for 5 min at 4˚C and the supernatant was transferred to the GC vials (1.8 mL). Next, 0.8 mL of the sample was combined with 0.2 mL of 25% phosphoric acid and 0.2 mL of internal standard solution. Standards (for acetic, propionic, isobutyric, butyric, isovaleric, valeric, and caproic acids) were prepared by combining 1 mL of standard solution and 0.2 mL of internal standard solution. The GC analysis was performed using the column Stabilwax-DA 30 meter (Restek Corp), the head pressure of 7.5 psi, split vent flow of 20 mL/minute, and injector temperature of 170ºC.

4.2.3 RNA extraction and sequencing

Total RNA was extracted from rumen samples using the TRIzol protocol based on the acid guanidinium-phenol-chloroform method with modifications reported previously

(Fuyong Li and Le Luo Guan 2017). Approximately 200 mg of rumen sample was subjected to RNA extraction with the addition of 1.5 ml of TRIzol reagent (pH: 4.6; Invitrogen, Carlsbad, CA, USA), followed by 0.4 ml of chloroform (pH: 7.0), 0.3 ml of isopropanol (pH: 7.0), and 0.3 ml of high salt solution (pH: 8.0) (1.2 M sodium acetate, 0.8 M NaCl) for the extraction protocol. The yield and integrity of the RNA samples were determined using a Qubit 2.0 fluorimeter (Invitrogen, Carlsbad, CA, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA samples were subjected to downstream RNA-sequencing only if they exhibited RNA with integrity number (RIN) higher than 7.0. Total RNA (100 ng) of each sample was used for library construction using the TruSeq RNA sample prep v2 LS kit (Illumina, San Diego, CA, USA) without the mRNA enrichment step. The quality of libraries was assessed using Agilent 2200 TapeStation (Agilent Technologies) and Qubit 2.0 fluorimeter (Invitrogen). Finally, cDNA fragments (~140 bp) were paired-end (2 X 100 bp) sequenced using an Illumina HiSeq 2000 system at the McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada).

4.2.4 Bioinformatics analysis

Taxonomic annotation of metatranscriptomes was obtained through the software Kraken (Wood and Salzberg 2014) as implemented in a pipeline developed by Neves et al. (2017). Metatranscriptome functional annotations were identified using ShotMAP as described by Nayfach et al. (2015). In summary, all fastq-formatted sequences were firstly analyzed through FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) for quality control, and then the software Trimmomatic (version 0.32) (Bolger et al. 2014) was used to trim residual artificial sequences, cut bases with quality scores below 20, and remove reads shorter than 50 bp. After the pre-processing steps described previously, SortMeRNA (version 1.9) (Kopylova et al. 2012) was used to sort the filtered reads into fragments of mRNA for microbial classification. Prior to performing the microbial classification, a Kraken standard database was built based on all complete genomes of bacteria downloaded from NCBI (RefSeq) plus complete genomes from organisms isolated from the rumen or from ruminant feces or saliva deposited in the Hungate1000 project (JGI's IMG database)) (Neves et al. 2017). Then, each pair of mRNA sequences

was assembled by MEGAHIT (Li et al. 2015), with the resulting contigs being assigned by Kraken (through k-mer discrimination) to the lowest common ancestor in the customized standard database for bacterial classification. Full taxonomic names associated with each classified sequence (*kraken* option *--preload*) and standard ranks (from domain to species) for each taxon were provided by the options *kraken-translate* and *kraken-mpa-report* (Neves et al. 2017).

Finally, the functions (microbial gene families) were obtained by mapping the assembled reads to the KEGG (Kanehisa et al. 2004) Orthology database using ShotMAP (Nayfach *et al.*, 2015), which is an algorithm based on the aligner RAPSearch2. Bit-score cutoffs (option *--class-score* 200) for matching contigs to the protein families were selected based on the average read length of each sample as described by Nayfach et al. (2015). The KEGG Orthology (KO) database was chosen because it annotates a large number of bacteria, including many species observed in the rumen microbiome, and covers a wide range of gene families, including metabolic enzymes, and signaling proteins. Average genome size (ags) was not estimated in this study (option *--ags-method* none) because we used mRNA data to annotate the microbial genes (Nayfach *et al.*, 2015).

4.2.5 Statistical analysis

To avoid sparsely observed counts, which tend to introduce noise in the analysis, we considered bacterial profiles and microbial functions with a relative abundance > 0.05% prevalent in at least 50% of the samples (12 out 24). Next, the filtered datasets were normalized by total sum scaling (TSS) (dividing each taxon/function count by the total number of counts in each individual sample to account for uneven sequencing depths across samples) and then transformed by centered log ratio (CLR) to project the data from a simplex to a Euclidean space (Aitchison 1982) as described in the *mixMC* multivariate statistical framework (Cao et al. 2016) (R package *mixOmics*) (Rohart *et al.*, 2017). Then, sparse partial least square discriminant analysis (sPLS-DA) (Cao et al. 2016) was applied to identify microbial signatures in Black Angus and Red Angus cattle while handling the sparsity of microbiome datasets and the sampling repeatedly performed on the same subjects (0, 80, 100, and 180 days). sPLS-DA includes a multilevel

decomposition approach that enables the detection of subtle differences in the signatures when high inter-subject variability is present due to sampling performed repeatedly on the same subjects, and the samples collected from four-time points in each breed were combined for sPLS-DA analysis. The ANCOM procedure (Mandal *et al.*, 2015), which also uses log-ratio transformations (Aitchison 1982), was performed to test whether the features of the filtered datasets were associated with FCR. The zero counts were replaced by ANCOM for an arbitrary pseudo count value of 0.001 and the normalized log-ratio of the abundance of each taxon/function was compared to the abundance of all remaining taxa/functions one at a time. Thereafter, ANOVA was calculated on the log-ratio matrix to find differentially abundant bacterial species and microbial functions between Black and Red Angus, adjusted for FCR and the time points. In ANCOM, the *p*-values were corrected for false discovery rate (FDR) using the Benjamini-Hochberg algorithm, and a threshold of FDR lower than 0.05 was applied to determine statistical significance. Finally, regularized canonical correlation (rCC) analysis (Gonzalez et al. 2008) was used to investigate the relationship between the VFA and bacterial species. Before running rCC analysis, the microbial data were normalized by TSS and then transformed by CLR, as described previously. The estimation of regularization parameters (λ1 and λ2) and canonical correlations were calculated using cross-validation procedures (Gonzalez et al. 2008). All statistical procedures and figures were done in R 3.4.2 (R Core Team, 2017) and Python 3.6.0.

4.2.6 Data repository resources

The datasets analyzed in this study were submitted to NCBI Sequence Read Archive (SRA) under the accession number PRJNA496209.

4.3 Results

4.3.1 Overview of the classification of active bacterial taxa

Assembly of mRNA reads resulting from total RNA sequencing of rumen samples generated a total of 7,330 contigs (with an average extension of 473.31 ± 30.34 bp and N50 of 434 ± 31.04 bp), which were further classified using a taxonomic assignment approach developed by our group. Approximately 53% of the mRNA reads were mapped to the contigs, indicating that they represented a significant proportion of the examined

rumen metatranscriptome. From the total number of phyla (n = 19) identified in the rumen of both breeds, Bacteroidetes (46.3%), Firmicutes (22.7%), Proteobacteria (14%), Spirochaetes (5%), Verrucomicrobia (2.3%), Tenericutes (2.2%), Actinobacteria (2%), and Fibrobacteres (1.4%) were the most abundant taxa, accounting for approximately 96% of the bacterial population in the rumen of all bulls (Figure 4.1A). A total of 109 families were detected and the most abundant phylotypes were assigned to Prevotellaceae (36.2%), Lachnospiraceae (6.0%), Ruminococcaceae (5.7%), Spirochaetaceae (5.3%), Bacteroidaceae (3.0%), Porphyromonadaceae (2.6%), Enterobacteriaceae (2.1%), Clostridiaceae (2.0%), and Fibrobacteraceae (1.7%) (Figure 4.1A). Our analysis also revealed a total of 114 genera, with *Prevotella* (41.5%), *Treponema* (4.6%), *Bacteroides* (3.5%), *Ruminoccocus* (3.5%), *Ruminiclostridium* (2.7%), *Butyrivibrio* (2.4%), *Clostridium* (2.2%), and *Fibrobacteres* (2.0%) being the most abundant in the rumen of all bulls (Figure 4.1A). Moreover, 114 bacterial species were classified in all samples, including *Prevotella ruminicola* (35.8%), *Butyrivibrio proteoclasticus* (3.3%), *Treponema succinifaciens* (2.9%), *Fibrobacter succinogenes* (1.8%), *Ruminiclostridium* sp KB18 (2.0%) and *Ruminococcus albus* (1.1%) (Figure 4.1A). A complete list of all bacterial phylotypes is provided in Supplementary Table S1.

4.3.2 Overview of the active microbial functions

To investigate the functional potential of the rumen microbiota, ShotMAP was used to survey which microbial functions (microbial gene families) were encoded in the microbiome by mapping the mRNA reads to the KEGG Orthology (KO) database. It was identified 109 active microbial functions in the rumen microbiome of the bulls and most of them were associated with ribosome, Calvin cycle, reductive citrate cycle, gluconeogenesis, glycolysis, and citrate cycle modules (Figure 4.1B). Among the various functions encoded by the bacteriome, pyruvate - orthophosphate dikinase (K01006, 6.8%), pyruvate-ferredoxin/flavodoxin oxidoreductase (K03737, 3.3%), DNA-directed RNA polymerase subunit beta (KO3046, 3.2%), glyceraldehyde 3-phosphate dehydrogenase (K00134, 2.7%), small subunit ribosomal protein S1 (K02945, 1.7%), phosphoenolpyruvate carboxykinase (ATP) (K01610, 1.4%), succinate dehydrogenase / fumarate reductase, flavoprotein subunit (K00239, 1.4%), and large subunit ribosomal

protein L1 (K02863, 1.2%) were the most abundant pathways within their respective modules (Figure 4.1B). A complete list of all microbial functions identified in this study is provided in Supplementary Table S2.

4.3.3 Identification of active microbial and functional signatures across breeds

To gain insight into the bacterial and functional signatures characterizing the rumen microbiome in Black and Red Angus, sPLS-DA was used to identify bacteria species and functions characterizing each breed. Based on the best classification error rate determined through cross-validation approaches two components were selected for the bacterial signature and only one component for the functional signature (Figures S1 and S2). Additionally, comparisons between CLR transformations of counts data with relative abundance data were performed. The results showed the relative abundance data generated only one component as the classification error rate increased after the addition of a second component (Figures S3 and S4). It was evident that some rumen bacterial species of the Black Angus signature overlapped in component 1 in both datasets (CLR vs. Relative Abundance) (Figure 4.2B and Figure S4). However, the CLR transformed data allowed the addition of a second component, improving the outcomes considerably by providing further information for the bacterial signature of Red Angus cattle (Figure 4.2B).

Following the CLR transformation procedures, a clear separation in bacteria species and functions differentiating the rumen microbiome of Black Angus from Red Angus cattle was identified (Figures 4.2A and 4.3A). Overall, 80% of the bacterial signature selected in component 1 of the sPLS-DA characterized the rumen microbiome of Black Angus (Figures 4.2B and 4.2C), and this bacterial signature included members of the families Chitinophagaceae (*Chitinophaga pinensis),* Clostridiaceae (*Clostridium stercorarium, Clostridium cellulosi,* and *Clostridium clariflavum*), Ruminococcaceae (*Ruminoccocus albus* and *Ruminococcus bicirculans*), Bacteroidaceae (*Bacteroides salanitronis*), Porphyromonadaceae (*Parabacteroides distasonis*), and Paludibacteraceae (*Paludibacter propionicigenes).* On the other hand, the component 2 of the sPLS-DA revealed that 60% of the bacterial signature was associated with Red Angus, and this bacterial signature was comprised of the following species: *Oscillibacter*

valericigenes (Oscillospiraceae), *Flavonifractor plautii* (Clostridiales), *Acidaminococcus fermentans* (Acidaminococcaceae), *Paenibacillus graminis* (Paenibacillaceae), and *Prevotella ruminicola* (Prevotellaceae) (Figures 4.2B and 4.2C).

For the microbial functional signature, the rumen microbiome of Black Angus displayed a diverse set of pathways including genetic information processing (e.g., K02878, K02992), environmental information processing/membrane transport (e.g., K10108), elongation factors for protein biosynthesis (e.g., K02355) and amino acid metabolism (e.g., K01740) (Figure 4.3B and 4.3C). However, the rumen microbiome of Red bulls was mostly enriched with pathways related to carbohydrate metabolism (e.g., K00895, K01785) (Figure 4.3B and 4.3C). A complete list of bacterial species and functions characterizing the rumen microbiome of Black and Red Angus (per component of the sPLS-DA) is presented in Supplementary Table S3.

4.3.4 Relationship between active bacteria and volatile fatty acids

rCC analysis was implemented to investigate interactions between bacterial taxa and VFAs (Figure 4.4A), as well as to identify interactions among different bacterial species (Figure 4.4B). Associations between bacteria and VFAs revealed a positive correlation between *R. albus* and total VFA, acetic and propionic acids (Figure 4.4A). Additionally, a positive correlation was observed between *P. ruminicola* and *S. ruminantium* and propionic, butyric and valeric acids, as well as between *O. valericigenes* and propionic, butyric, and valeric acids (Figure 4.4A). On the other hand, negative correlations were detected between *C. pinensis* and *A. fermentans* and total VFA, acetic, propionic, butyric and valeric acids, as well as between *C. stercorarium* and propionic, butyric and valeric acids (Figure 4.4A). Within the bacterial communities, *P. ruminicola, S. ruminantium* and *R. albus* were negatively correlated with *A. fermentans* (Figures 4.4B), while *C. stercorarium* and *C. pinensis* were positively correlated with each other and with *R. bicirculans, Ruminiclostridium* sp KB18 and *C. cellulosi* (Figure 4.4B).

4.3.5 Relationship between active bacteria and feed efficiency

To evaluate the linkage among bacteria, functions and feed efficiency (measured as FCR), ANCOM was used to identify taxa and functions differentially ($P < 0.05$) abundant

between breeds, adjusted for FCR and the time points. Of the 114 bacteria species classified in all samples, it was found that only two species were associated ($P < 0.05$) with FCR according to the breeds (Figure 4.5). *Chitinophaga pinensis* showed a relative abundance 2.5 times higher ($P < 0.05$) in the rumen of Black Angus compared to Red Angus associated with FCR and the four-time points (Figure 4.5A). *Clostridium stercorarium* also exhibited a higher ($P < 0.05$) relative abundance in the rumen of Black Angus compared to Red Angus associated with FCR, but only on the days 0 and 180 of the experimental period (Figures 4.5B). No link between the functional signatures detected by ShotMAP and FCR was found in the rumen of Black and Red Angus cattle.

4.4 Discussion

Until now only a few studies have addressed the question of the breed effect on the rumen microbiome in beef cattle (Roehe et al. 2016, Guan et al. 2008, Z. P. Li et al. 2016). First, PCR-denaturating gradient gel electrophoresis revealed that the rumen microbiome of beef steers exhibiting different feed efficiencies (measured as residual feed intake) clustered according to their breeds (Angus, Charolais, and Hereford-Angus), suggesting that host breed may play a role in shaping the structure of the rumen microbiota (Guan et al. 2008). Second, a metagenomic analysis of the rumen microbiome in Limousin- and Aberdeen Angus-sired cattle showed that the abundance of microbial genes involved in methanogenesis and feed efficiency (measured as FCR) could be used to predict host metabolism, performance, and behavior (Roehe et al. 2016). Third, a metagenomic analysis of the rumen microbiome in crosses between sika deer and elk reported that the rumen microbiota in the hybrids differed from their parents, suggesting a significant effect of host genetics on the rumen microbiome likely caused by vertical transmission of the maternal microbiota (Z. P. Li et al. 2016). Although those three investigations suggested a connection of the host genetics with the rumen microbiome, those studies used DNA-based methods, which do not give a clear assessment of gene expression within the rumen microbiota (Franzosa et al. 2014). More recently, Li et al. (2019a) compared rumen metagenomic and metatranscriptomic datasets of three breeds of beef cattle (Angus, Charolais, and Kinsella composite hybrid) and revealed that the rumen metatranscriptome represents the functional activities of microbes and is more useful than

metagenomics to show the associations between rumen microbes and host phenotypical traits.

However, previous studies using rumen-based metagenomics or metatranscriptomics (Roehe et al. 2016, Guan et al. 2008, Z. P. Li et al. 2016) have ignored that sequencing data hold a compositional structure because of necessary relative proportion transformation (Cao et al. 2016, Mandal et al. 2015, Gloor et al. 2016, Aitchison 1982). Thus, the traditional statistical methods (including Pearson correlations, PCA, PCoA, GLM, partial least square analysis, ANOVA, linear regression) cannot be applied directly in the analysis of relative abundance data because the independence assumption between predictor variables is not met, which may result in spurious results (Cao et al. 2016, Aitchison 1982, Lovell et al. 2015, Ban et al. 2015, Kurtz et al. 2014). A recent study (Weiss et al. 2017) evaluating seven statistical methods for differential abundance testing suggested that a novel methodology (ANCOM) based on log-ratio transformations of count data, as defined by Aitchison (1982), was the most effective approach to control false discovery rates. Similarly, Lê Cao and colleagues developed sPLS-DA based on centered log ratio transformations of count data to identify microbial signatures from diverse microbiomes (Rohart et al. 2017). Here, the transformations of counts data implemented in these statistical methods provided more reliable outcomes for the bacterial signatures than the relative abundance data and enabled a further understanding of the relationships that exist between breed-specific bacterial signatures, host phenotype (Black vs. Red Angus) and feed efficiency at the RNA-level.

Microbial signatures (or biomarkers) reflect correspondences between microbial taxa and functions and are widely used to predict host phenotypes in human disease states and forensic diagnostics (Knights et al. 2011, San-Juan-Vergara et al. 2018). In general, it is unlikely that taxa that comprise the core rumen microbiome (including *Prevotella*, *Treponema*, *Ruminoccocus*, *Fibrobacter*, and *Butyrivibrio*) will constitute breed-specific bacterial signatures, since these taxa are prevalent in all ruminants (bovines, camelids, caprids, cervids) regardless of sex, age, and breed (Henderson et al. 2015). Nonetheless, our results showed a breed effect on the core rumen microbiome for *Prevotella ruminicola*, since this bacterium exhibited unexpectedly a higher abundance in

the rumen of Red Angus compared to Black Angus fed the same diet. As both breeds were raised under the same nutritional regimen, it is probable that *P. ruminicola* was more active in Red Angus to occupy specific ecological niches within the rumen, like the degradation of hemicellulose and pectin (Russell and Rychlik 2001). Thus, we speculate that if *P. ruminicola* is part of the microbial signature predicting the Red Angus phenotype, then this biological signature may be used as a target for rumen manipulation purposes in that specific breed. However, further investigations on a larger set of animals should be conducted to differentiate the rumen microbiome of Red from Black Angus raised in the same or in different geographical/environmental conditions to assess the robustness of *P. ruminicola* as a differentiator of these two closely related breeds.

Moreover, we observed that Red Angus showed a microbial signature comprised of bacterial phylotypes (e.g., *Oscillibacter valericigenes* and *Flavonifractor plautii*) that were not yet described in the rumen microbiota from previous studies. In the gut of pigs, *O. valericigenes* presented a positive correlation with valerate and butyrate production (Pajarillo et al. 2015), which are short-chain fatty acids associated with improved feed efficiency in beef cattle (Guan et al. 2008). In this study, *O. valericigenes* also showed a positive correlation with valeric and butyric acids, suggesting that these microbes played a role in the ruminal kinetics of Red Angus cattle, especially in the metabolism of valerate and butyrate. Another microbe we detected in the bacterial signature of Red Angus was *Acidaminococcus fermentans*. At the metagenomic level, a recent study recovered and assembled the genome of *A. fermentans* from rumen samples collected from Aberdeen Angus, Limousin, Charolais, and Luing (Stewart et al. 2018), confirming that this bacterium is a ubiquitous inhabitant of the rumen microbiota in beef cattle. Possible activities that *A. fermentans* perform in the rumen include the utilization of citrate as an energy source to produce hydrogen and hydrogen sulfide and the decrease in the accumulation of tricarballylate (a toxic end-product of ruminal fermentation) by oxidizing *trans*-aconitate (Cook et al. 1994, Wallace et al. 2015, Stewart et al. 2018), demonstrating that *A. fermentans* plays important roles in the Red Angus bacterial signature.

In contrast, the Black Angus bacterial signature exhibited a larger number of bacteria species responsible for the degradation of lignocellulosic biomass. One example is *Chitinophaga pinensis,* which has been reported to encode a diverse array of glycoside hydrolases (e.g., the GH27 enzyme *CpArap27*) capable of utilizing plant biomass-derived carbohydrates (Larsbrink et al. 2017a, Larsbrink et al. 2017b). Other instances of lignocellulose-degrading microbes we found in Black Angus bacterial signature included *Ruminoccocus albus* and *Clostridium stercorarium*, which are fibrolytic species actively involved in the digestion of cellulose/xylan (Christopherson et al. 2014) and hemicellulose (Schellenberg et al. 2014), respectively. In addition to showing a higher abundance in the rumen of Black Angus compared to Red Angus, *C. pinensis* and *C. stercorarium* were the only species directly connected with feed efficiency measurements (FCR) taken across the feeding period, confirming the beneficial effects of these two species on the deconstruction of plant cell wall fibers in forage-fed cattle (Christopherson et al. 2014, Larsbrink et al. 2017a). With its limited sample size, however, our study lacked the power necessary to find differences in feed efficiency (FCR) between Black and Red Angus (Figure S5). Although FCR measurements did not differ significantly between the two breeds, evidence for superior performance of Black Angus compared to Red Angus has been described previously (McLean and Schmutz 2009). However, other studies showed that Red Angus had a better gain to feed ratios than Black Angus, demonstrating that differences in feeding behavior exist between these two genetically similar breeds (Wolfger et al. 2016). Thus, it is necessary to further investigate whether *C. pinensis* and *C. stercorarium* would contribute to growth performance in Angus cattle exhibiting different feed efficiencies (e.g., high vs. low FCR) under various feeding regimens to validate our findings. Although three animals per breed may not confer the necessary power to detect differences in FCR with four-time points per animal in the current study, the variations in those two rumen bacteria associated with FCR are valid. However, more animals should be included in the future to support our findings (Figure S5).

Understanding the functions of the rumen microbiome is crucial to the development of technologies and practices that support efficient global food production from ruminants (Seshadri et al. 2018). The identification of functional signatures in the current study provided evidence that the functions performed by the rumen microbiota

depend on breed. The rumen functional signature found in Red Angus was to a large extent related to carbohydrate metabolism, corroborating the findings reported in previous metagenomic studies that showed a high abundance of carbohydrate-related functions in cattle fed forage diets (Wang et al. 2013). In this context, it was not a surprise to detect active pathways involved mainly in the fermentation of glucose to acetate from Acetyl-CoA via the Embden-Meyerhof-Parnas (EMP) pathway (Russell and Rychlik 2001, Gottschalk 1986). However, it was found an enzyme in Red Angus called phosphofructokinase [EC:2.7.1.90, K00895] that can form acetate through an atypical (incomplete) EMP pathway. A recent study has provided evidence that atypical EMP pathways are encoded in the genome of four rumen bacterial species (including *P. ruminicola*) (Hackmann et al. 2017), which showed a higher abundance in the Red Angus rumen microbiome as discussed previously. According to Hackmann *et al.* (2017), those bacteria could convert glucose into pyruvate through the supply of phosphofructokinases [EC:2.7.1.90, K00895] rather than from Acetyl-CoA.

Unlike Red Angus cattle, the functional signature in Black Angus exhibited an increase in microbial functions associated with protein synthesis during the assembly of the bacterial ribosome (e.g., ribosomal protein L16 [K02878] and ribosomal protein S7 [K02992]). The mechanisms of action of proteins L16 have been described in *Escherichia coli* and *Thermus thermophilus* HB8 (Anikaev et al. 2016). These studies have shown that mutations in those proteins lead to a significant reduction in cell growth and a decrease in their translation apparatus. Mutations in the protein L16, which participates in the formation of functional sites of the 50S ribosomal subunit (Anikaev et al. 2016), can activate the binding sites for antibiotics in the internal cavity of the bacterial ribosome. In a similar fashion, mutations in the binding sites of conserved regions (30S subunit) of the bacterial RNA in protein S7 can destabilize the correct three-dimensional structure of the assembled ribosome (Wimberly et al. 1997). Although our findings showed that the proteins L16/S7 were more relevant to Black Angus than to Red Angus, further studies exploring relationships between proteins L16/S7 and bacterial growth dynamics in other beef cattle breeds are needed to validate our results. Future directions may also include studying the interactions between proteins L16/S7 and RNA-biding sites with the goal of

designing antibiotics with novel structures and modes of action to lower the increasing bacterial resistance to commonly used antimicrobials in the beef cattle industry.

4.5 Conclusions

This study described the active microbial and functional signatures of the rumen microbiome in two closely related breeds using total RNA-seq-based metatranscriptomics. Our results showed that Black and Red Angus exhibited differences in microbial signatures at the functional and compositional levels, indicating that the breed influences the structure of the rumen microbiome and its function. It is noticeable that the bacterial signature in Black Angus was characterized by an increase in fibrolytic species such as *Chitinophaga pinensis, Clostridium stercorarium,* and *Ruminococcus albus*, whereas the bacterial signature in Red Angus was composed of poorly characterized species (e.g., *Oscillibacter valericigenes* and *Flavonifractor plautii*). Interestingly, FCR was associated with a specific bacterial signature depending on the breed type, suggesting that there is a connection among feed efficiency, active rumen microbiota, and the genetic makeup of the host. In this context, *C. pinensis* and *C. stercorarium* were the only species associated with FCR, and this result emphasizes the important role of these species in the feed conversion efficiency of forage-fed bulls. Moreover, our results showed specific functional signatures characterizing each breed (separately), demonstrating that different strategies to modify the rumen microbial pathways need to be designed for Red Angus as compared to Black Angus. While the first showed microbial functions enriched mainly from carbohydrate metabolism, the latter exhibited a more diverse set of pathways, with emphasis on the ribosomal proteins L16 and S7. It is important to mention that the statistical methods employed in this study can also be applied to investigate rumen archaea, fungi, and protozoa in addition to bacteria as reported in the current study. However, the lack of genome sequences representatives of rumen archaeal, fungal, and protozoal populations (especially in the publicly available database - e.g., NCBI) does not allow accurate taxonomic assessment at species level for these organisms by Kraken. Thus, future studies involving the rumen microbiome as a whole are warranted to enhance our knowledge of the symbiotic relationship that exists among rumen microorganisms and how they may differ between these two genetically similar breeds.

Regardless, our findings highlight how we can use signatures of bacterial taxa and their functions to harness the full potential of the rumen microbiome in Angus cattle. Furthermore, the results of our study can be used as a reference for future investigations related to the manipulation of the ruminal function and selection of beef cattle with high feed efficiency based on specific microbial signatures.

4.6 Literature cited

Aitchison, J. (1982) 'The statistical analysis of compositional data', *Journal of the Royal Statistical Society Series B-Methodological,* 44(2), 139-177.

Anikaev, A., Isaev, A., Korobeinikova, A., Garber, M. and Gongadze, G. (2016) 'Role of protein L25 and its contact with protein L16 in maintaining the active state of Escherichia coli ribosomes in vivo', *Biochemistry (00062979),* 81(1), 19-27.

Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. and Segata, N. (2015) 'Compact graphical representation of phylogenetic data and metadata with GraPhlAn', *PeerJ,* 3.

Ban, Y., Jiang, H. and An, L. (2015) '- Investigating microbial co-occurrence patterns based on metagenomic compositional data', - 31(- 20), - 3329.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics (Oxford, England),* 30(15), 2114-2120.

Cao, K. A. L., Costello, M. E., Lakis, V. A., Bartolo, F., Chua, X. Y., Brazeilles, R. and Rondeau, P. (2016) 'MixMC: A multivariate statistical framework to gain insight into microbial communities', *Plos One,* 11(8).

Christopherson, M. R., Dawson, J. A., Stevenson, D. M., Cunningham, A. C., Bramhacharya, S., Weimer, P. J., Kendziorski, C. and Suen, G. (2014) 'Unique aspects of fiber degradation by the ruminal ethanologen Ruminococcus albus 7 revealed by physiological and transcriptomic analysis', *BMC Genomics,* 15.

Cook, G. M., Rainey, F. A., Chen, G., Stackebrandt, E. and Russell, J. B. (1994) 'Emendation of the description of Acidaminococcus fermentans, a trans- aconitate- and citrate-oxidizing bacterium', *International Journal Of Systematic Bacteriology,* 44(3), 576-578.

Creevey, C. J., Kelly, W. J., Henderson, G. and Leahy, S. C. (2014) 'Determining the culturability of the rumen bacterial microbiome', *Microbial Biotechnology,* 7(5), 467-479.

Crews, D. H., Jr., Shannon, N. H., Genswein, B. M. A., Crews, R. E., Johnson, C. M. and Kendrick, B. A (2005) 'Genetic parameters for net feed efficiency of beef cattle measured during postweaning growing versus finishing periods', *Proc. West Sec. Am. Soc. Anim. Sci,* 54(54), 125-128.

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., Giannoukos, G., Boylan, M. R., Ciulla, D., Gevers, D., Izard, J., Garrett, W. S., Chan, A. T. and Huttenhower, C. (2014) 'Relating the metatranscriptome and metagenome of the human gut', *Proceedings of the National Academy of Sciences,* 111(22), E2329-E2338.

Geishauser, T. (1993) 'An Instrument for collection and transfer of ruminal fluid and for administration of water soluble drugs in adult cattle', *The Bovine Practitioner,* (27), 38-42.
Gloor, G., Wu, J. R., Pawlowsky-Glahn, V. and Jose Egozcue, J. (2016) '- It's all relative: analyzing microbiome data as compositions', - 26(- 5), - 329.
Gonzalez, I., Dejean, S., Martin, P. G. P. and Baccini, A. (2008) 'CCA: An R package to extend canonical correlation analysis', *Journal of Statistical Software,* 23(12), 1-14.
Gottschalk, G. (1986) 'Bacterial metabolism',
Guan, L. L., Nkrumah, J. D., Basarab, J. A. and Moore, S. S. (2008) 'Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency', *Fems Microbiology Letters,* 288(1), 85-91.
Hackmann, T. J., Tao, J., Ngugi, D. K. and Firkins, J. L. (2017) 'Genomes of rumen bacteria encode atypical pathways for fermenting hexoses to short-chain fatty acids', *Environmental Microbiology,* 19(11), 4670-4683.
Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) 'The KEGG resource for deciphering the genome', *Nucleic Acids Research,* 32(Database issue), D277-D280.
Knights, D., Parfrey, L. W., Zaneveld, J., Lozupone, C. and Knight, R. (2011) 'Human-associated microbial signatures: examining their predictive value', *Cell Host & Microbe,* 10(4), 292-296.
Kopylova, E., Noé, L. and Touzet, H. (2012) 'SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data', *Bioinformatics (Oxford, England),* 28(24), 3211-3217.
Kurtz, Z. D., Mueller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J. and Bonneau, R. A. (2014) 'Sparse and compositionally robust inference of microbial ecological networks',
Larsbrink, J., Tuveng, T. R., Pope, P. B., Bulone, V., Eijsink, V. G. H., Brumer, H. and McKee, L. S. (2017a) 'Proteomic insights into mannan degradation and protein secretion by the forest floor bacterium Chitinophaga pinensis', *Journal of Proteomics,* 156, 63-74.
Larsbrink, J., Tuveng, T. R., Pope, P. B., Eijsink, V. G. H., Bulone, V., Brumer, H. and McKee, L. S. (2017b) 'Proteomic data on enzyme secretion and activity in the bacterium Chitinophaga pinensis', *Data in Brief,* 11, 484-490.
Lee, S., Sung, J., Lee, J. and Ko, G. (2011) 'Comparison of the gut microbiotas of healthy adult twins living in South Korea and the United States', *Applied and Environmental Microbiology,* 77(20), 7433-7437.
Li, D. H., Liu, C. M., Luo, R. B., Sadakane, K. and Lam, T. W. (2015) 'MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics,* 31(10), 1674-1676.
Li, F. and Guan, L. L. (2017) 'Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle', *Applied and Environmental Microbiology*.
Li, F., Hitch, T. C. A., Chen, Y., Creevey, C. J. and Guan, L. L. (2019) 'Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the

rumen microbiome and its associations with feed efficiency in beef cattle', *Microbiome,* 7(1), 6.
Li, Z. P., Wright, A. D. G., Si, H. Z., Wang, X. X., Qian, W. X., Zhang, Z. G. and Li, G. Y. (2016) 'Changes in the rumen microbiome and metabolites reveal the effect of host genetics on hybrid crosses', *Environmental Microbiology Reports,* 8(6), 1016-1023.
Lovell, D., Pawlowsky-Glahn, V., Egozcue, J. J., Marguerat, S. and Bähler, J. (2015) 'Proportionality: A Valid Alternative to Correlation for Relative Data', *Plos Computational Biology,* 11(3), 1-12.
Mandal, S., Van Treuren, W., White, R. A., Eggesbo, M., Knight, R. and Peddada, S. D. (2015) 'Analysis of composition of microbiomes: a novel method for studying microbial composition', *Microbial Ecology in Health and Disease,* 26, 27663-Article No.: 27663.
Márquez, G. C., Speidel, S. E., Enns, R. M. and Garrick, D. J. (2010) 'Genetic diversity and population structure of American Red Angus cattle', *Journal of Animal Science,* 88(1), 59-68.
McLean, K. L. and Schmutz, S. M. (2009) 'Associations of melanocortin 1 receptor genotype with growth and carcass traits in beef cattle', *Canadian Journal of Animal Science,* 89(3), 295-300.
Nayfach, S., Bradley, P. H., Wyman, S. K., Laurent, T. J., Williams, A., Pollard, K. S., Sharpton, T. J. and Eisen, J. A. (2015) 'Automated and accurate estimation of gene family abundance from shotgun metagenomes', *Plos Computational Biology,* 11(11), 29p.
Neves, A. L. A., Li, F., Ghoshal, B., McAllister, T. and Guan, L. L. (2017) 'Enhancing the resolution of rumen microbial classification from metatranscriptomic data using Kraken and Mothur', *Frontiers in Microbiology,* 8(2445).
Opdahl, L. J., Gonda, M. G. and St-Pierre, B. (2018) 'Identification of Uncultured Bacterial Species from Firmicutes, Bacteroidetes and CANDIDATUS Saccharibacteria as Candidate Cellulose Utilizers from the Rumen of Beef Cows', *Microorganisms,* 6(1).
Pajarillo, E. A. B., Chae, J. P., Balolong, M. P., Kim, H. B., Kang, D.-K. and Seo, K.-S. (2015) '- Characterization of the Fecal Microbial Communities of Duroc Pigs Using 16S rRNA Gene Pyrosequencing', - 28(- 4), - 591.
Roehe, R., Dewhurst, R. J., Duthie, C.-A., Rooke, J. A., McKain, N., Ross, D. W., Hyslop, J. J., Waterhouse, A., Freeman, T. C., Watson, M. and Wallace, R. J. (2016) 'Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance', *Plos Genetics,* 12(2).
Rohart, F., Gautier, B., Singh, A. and Lê Cao, K.-A. (2017) 'mixOmics: An R package for 'omics feature selection and multiple data integration', *Plos Computational Biology,* 13(11), 1-19.
Russell, J. B. and Rychlik, J. L. (2001) 'Factors that alter rumen microbial ecology', *Science,* 292(5519), 1119-1122.
San-Juan-Vergara, H., Peña, M., Cadena-Cruz, C., Diaz-Olmos, Y., Hurtado-Gómez, L., Vidal-Orjuela, O., Viasus, D., Cervantes-Acosta, G., Zurek, E., Portnoy, I., Vélez, J. I., Ajami, N. J., Petrosino, J. F., Mogollon, C., Hernández, D., Urruchurtu, I., Di-

Ruggiero, P., Sanchez-Sit, S., Guardo-García, E. and Torres, N. (2018) 'A Lachnospiraceae-dominated bacterial signature in the fecal microbiota of HIV-infected individuals from Colombia, South America', *Scientific Reports,* 8(1).

Schellenberg, J. J., Verbeke, T. J., Sparling, R., McQueen, P., Krokhin, O. V., Wilkins, J. A., Zhang, X., Alvare, G., Fristensky, B., Levin, D. B., Thallinger, G. G. and Henrissat, B. (2014) 'Enhanced whole genome sequence and annotation of Clostridium stercorarium DSM8532T using RNA-seq transcriptomics and high-throughput proteomics', *BMC Genomics,* 15(1).

Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Hadjithomas, M., Varghese, N. J., Paez-Espino, D., Perry, R., Henderson, G., Creevey, C. J., Terrapon, N., Lapebie, P., Drula, E., Lombard, V., Rubin, E., Kyrpides, N. C., Henrissat, B., Woyke, T., Ivanova, N. N. and Kelly, W. J. (2018) 'Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection', *Nature Biotechnology,* 36(4), 359-367.

Stewart, R. D., Warr, A., Watson, M., Auffret, M. D., Dewhurst, R. J., Roehe, R., Wiser, A. H., Press, M. O., Langford, K. W., Liachko, I., Snelling, T. J. and Walker, A. W. (2018) 'Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen', *Nature Communications,* 9(1).

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C.-A., Hyslop, J. J., Ross, D. W., Waterhouse, A., Watson, M. and Roehe, R. (2015) 'The rumen microbial metagenome associated with high methane production in cattle', *BMC Genomics,* 16, 1-14.

Wang, L., Hatem, A., Catalyurek, U. V., Morrison, M. and Yu, Z. (2013) 'Metagenomic Insights into the Carbohydrate-Active Enzymes Carried by the Microorganisms Adhering to Solid Digesta in the Rumen of Cows', *Plos One,* 8(11), 1-11.

Wang, Y. and McAllister, T. A. (2002) 'Rumen microbes, enzymes and feed digestion - A review', *Asian-Australas J Anim Sci,* 15(11), 1659-1676.

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vazquez-Baeza, Y., Birmingham, A., Hyde, E. R. and Knight, R. (2017) 'Normalization and microbial differential abundance strategies depend upon data characteristics', *Microbiome,* 5.

Wimberly, B. T., Ramakrishnan, V. and White, S. W. (1997) 'The structure of ribosomal protein S7 at 1.9 Å resolution reveals a β-hairpin motif that binds double-stranded nucleic acids', *Structure,* 5(9), 1187-1198.

Wolfger, B., Quinn, C., Torres, G. W., Taylor, M. and Orsel, K. (2016) 'Comparison of feeding behavior between black and red Angus feeder heifers', *Canadian Journal of Animal Science,* 96(3), 404-409.

Wood, D. E. and Salzberg, S. L. (2014) 'Kraken: ultrafast metagenomic sequence classification using exact alignments', *Genome Biology,* 15(3), R46-R46.

Zhou, M., Peng, Y.-J., Chen, Y., Klinger, C. M., Oba, M., Liu, J.-X. and Guan, L. L. (2018) 'Assessment of microbiome changes after rumen transfaunation: implications on improving feed efficiency in beef cattle', *Microbiome,* 6(1), 62-62.

4.7 Tables and figures

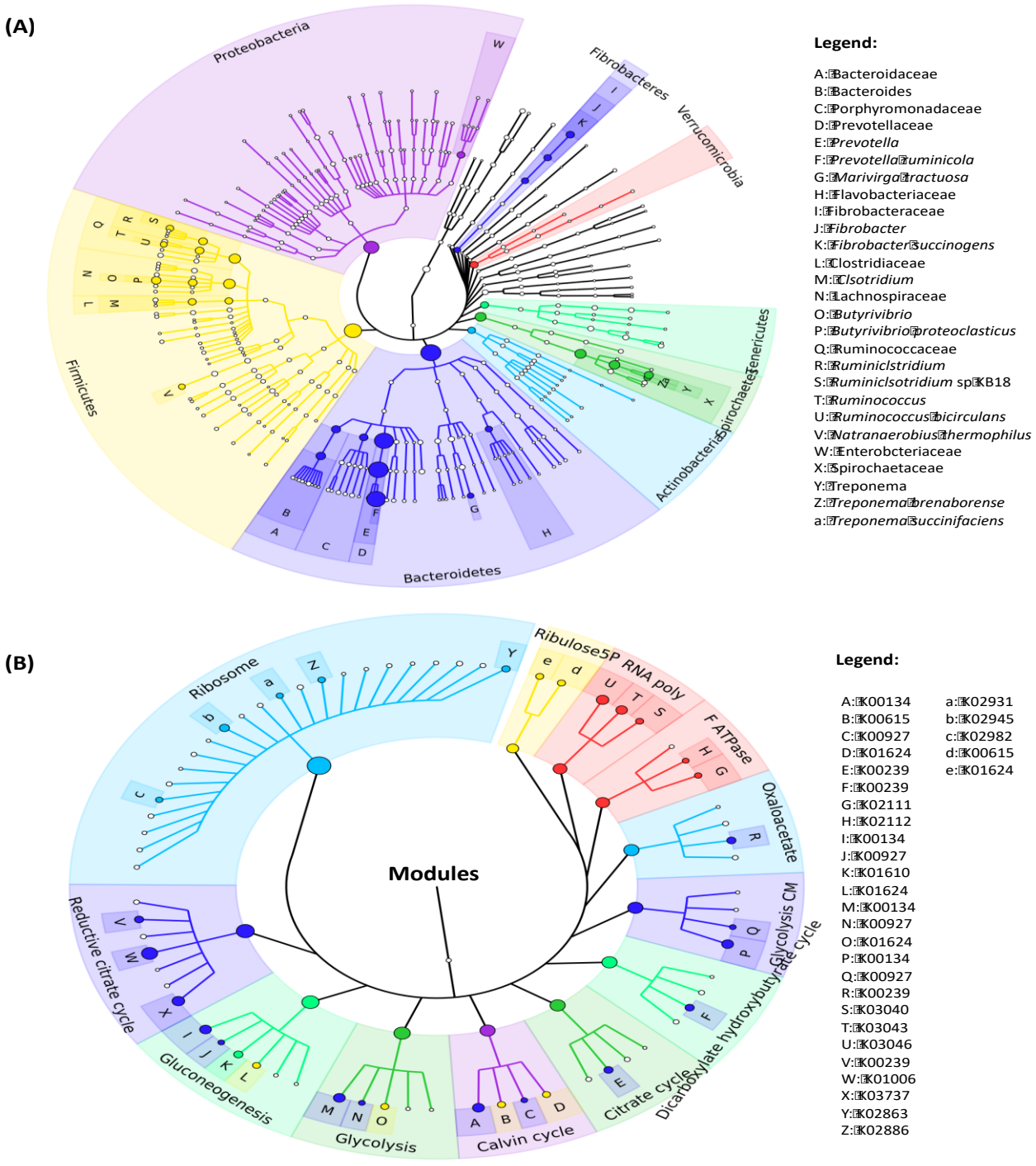


Figure 4.1 Taxonomic composition and microbial gene families obtained from metatranscriptome data in beef cattle. (A) Taxonomic cladogram showing the most

abundant detected taxa (relative abundance ≥ 0.05% in at least half of the samples). The six rings of the cladogram stand for phylum (innermost), class, order, family, genus, and species (outermost), respectively. The sizes of the circles indicate the mean average abundance of each taxon. (B) Cladogram showing the most abundant microbial functions (relative abundance 0.05% ≥ in at least half of the samples). Two rings of the cladogram stand for module (innermost) and microbial functions obtained from the KEEG database (outermost), respectively. The sizes of the circles indicate the mean average abundance of each function (microbial gene family). Cladograms were built using GraPhlan (Asnicar et al. 2015).

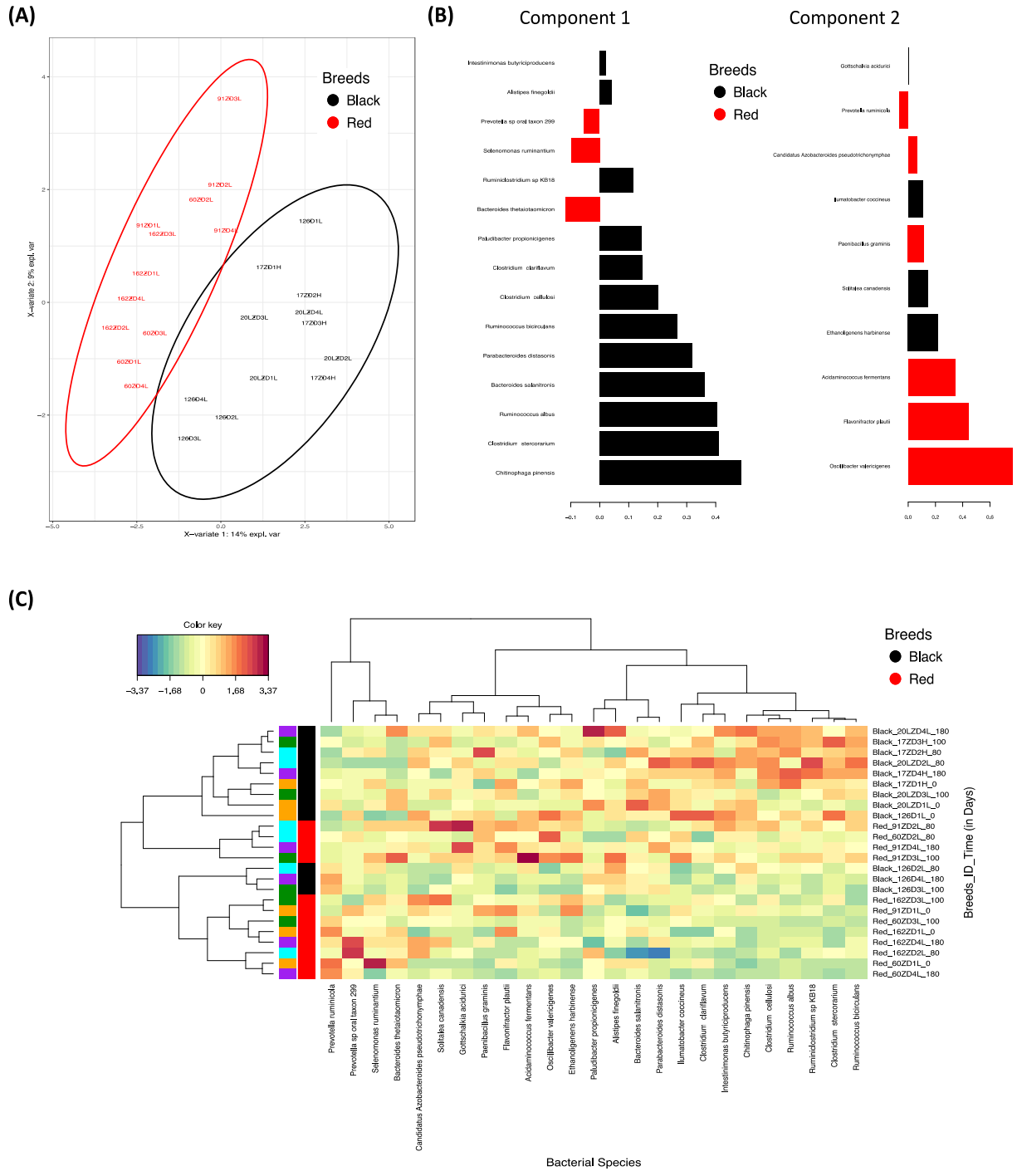


Figure 4.2 sPLS-DA results on rumen bacterial species in Black and Red Angus. (A) shows a sample plot on the first two sPLS-DA components with 95% confidence level

ellipse plots and (B) represents the contribution of each species selected on the first and second components, with length of the bar representing the importance of each species to the component (importance from bottom to top). Colors indicate the breed (Black vs. Red Angus) in which the species is most abundant. (C) represents a hierarchical clustering (Euclidean distance, Ward linkage) of the selected species from sPLS-DA results. In the heatmap legend, the red and blue colors indicate strong positive and negative correlations respectively, whereas yellow or green indicate weaker correlation values. On the left-hand side of the heatmap, clusters of the biological samples collected according to time are colored in orange, blue, dark green and violet to signify the time points 0, 80, 100 and 180 days, respectively.

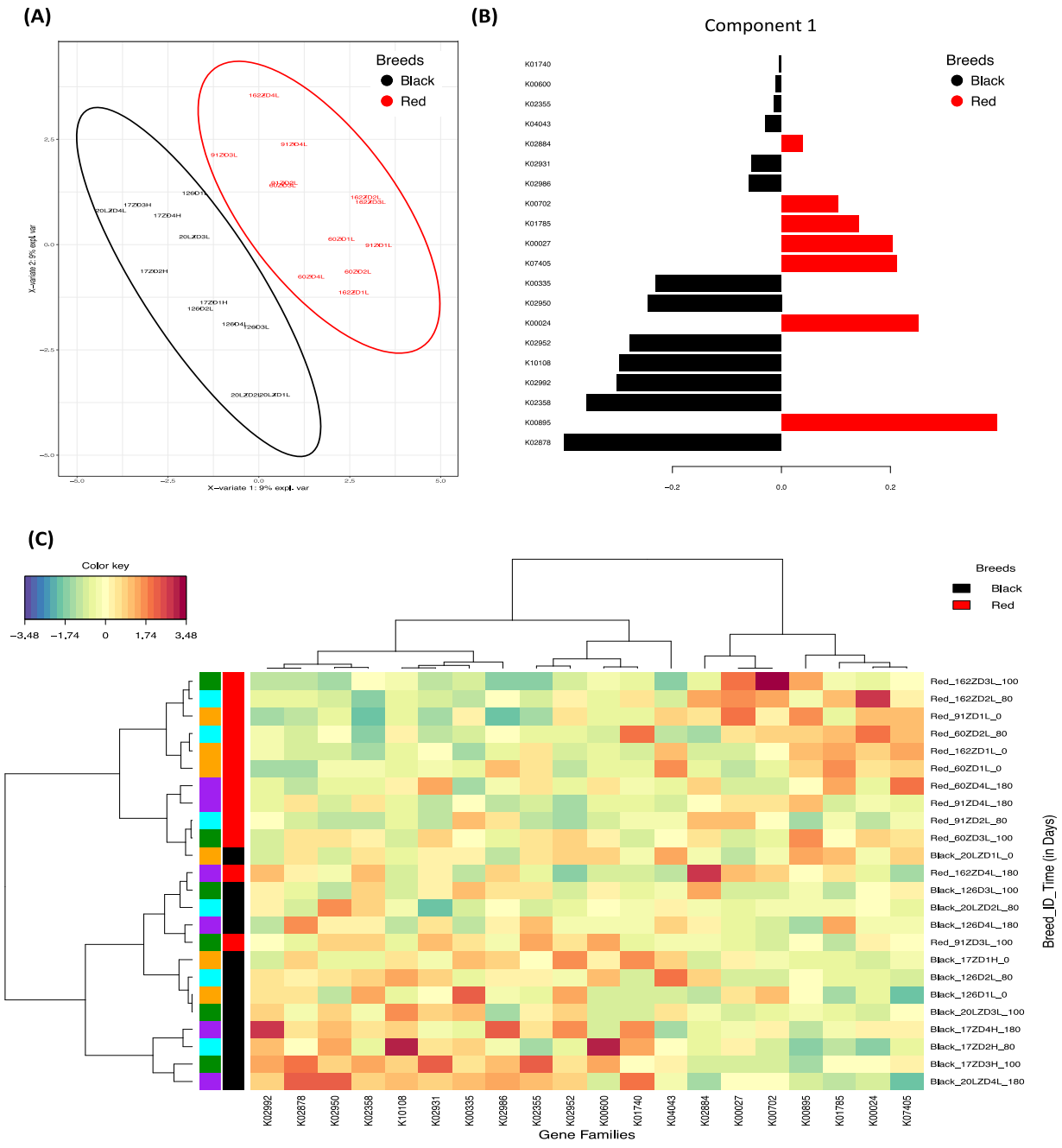
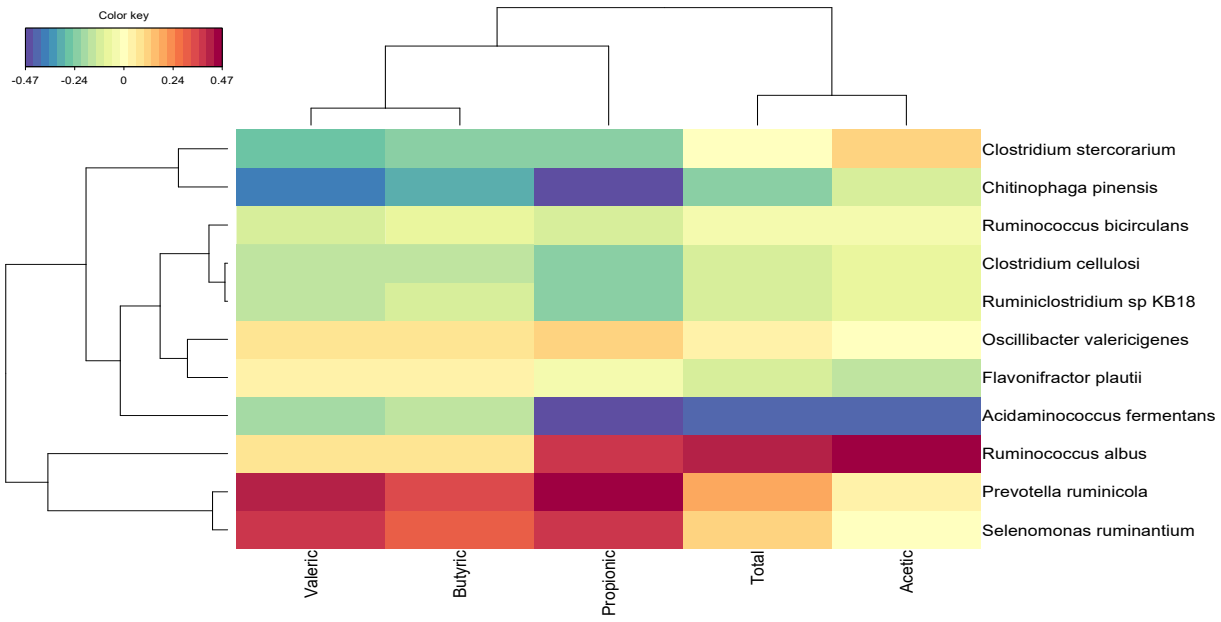

(A)
Breeds
Black
Red
X−variate 1: 9% expl. var
X−variate 2: 9% expl. var
(B)
Component 1
Breeds
Black
Red
K01740
K00600
K02355
K04043
K02884
K02931
K02986
K00702
K01785
K00027
K07405
K00335
K02950
K00024
K02952
K10108
K02992
K02358
K00895
K02878
(C)
Color key
−3.48 −1.74 0 1.74 3.48
Breeds
Black
Red
Red_162ZD3L_100
Red_162ZD2L_80
Red_91ZD1L_0
Red_60ZD2L_80
Red_162ZD1L_0
Red_60ZD1L_0
Red_60ZD4L_180
Red_91ZD4L_180
Red_91ZD2L_80
Red_60ZD3L_100
Black_20LZD1L_0
Red_162ZD4L_180
Black_126D3L_100
Black_20LZD2L_80
Black_126D4L_180
Red_91ZD3L_100
Black_17ZD1H_0
Black_126D2L_80
Black_126D1L_0
Black_20LZD3L_100
Black_17ZD4H_180
Black_17ZD2H_80
Black_17ZD3H_100
Black_20LZD4L_180
K02992
K02878
K02950
K02358
K10108
K02931
K00335
K02986
K02355
K02952
K00600
K01740
K04043
K02884
K00027
K00702
K00895
K01785
K00024
K07405
Gene Families
Breed_ID_Time (in Days)

Figure 4.3 sPLS-DA results on rumen microbial gene families in Black and Red Angus. (A) shows a sample plot on the first sPLS-DA component with 95% confidence level ellipse plots and (B) represents the contribution of each function (microbial gene family) selected on the first component, with length of the bar representing importance of each microbial gene family to the component (importance from bottom to top). Colors indicate the breed (Black vs. Red Angus) in which the gene family is most abundant. (C) represents the hierarchical clustering (Euclidean distance, Ward linkage) of the selected gene families from sPLS-DA results. In the heatmap legend, the red and blue colors indicate strong positive and negative correlations respectively, whereas yellow or green indicate weaker correlation values. On the left-hand side of the heatmap, clusters of the biological samples collected according to time are colored in orange, blue, dark green and violet to signify the time points 0, 80, 100 and 180 days, respectively.

(A)


Color key
-0.47
-0.24
0
0.24
0.47
Clostridium stercorarium
Chitinophaga pinensis
Ruminococcus bicirculans
Clostridium cellulosi
Ruminiclostridium sp KB18
Oscillibacter valericigenes
Flavonifractor plautii
Acidaminococcus fermentans
Ruminococcus albus
Prevotella ruminicola
Selenomonas ruminantium
Valeric
Butyric
Propionic
Total
Acetic

(B)

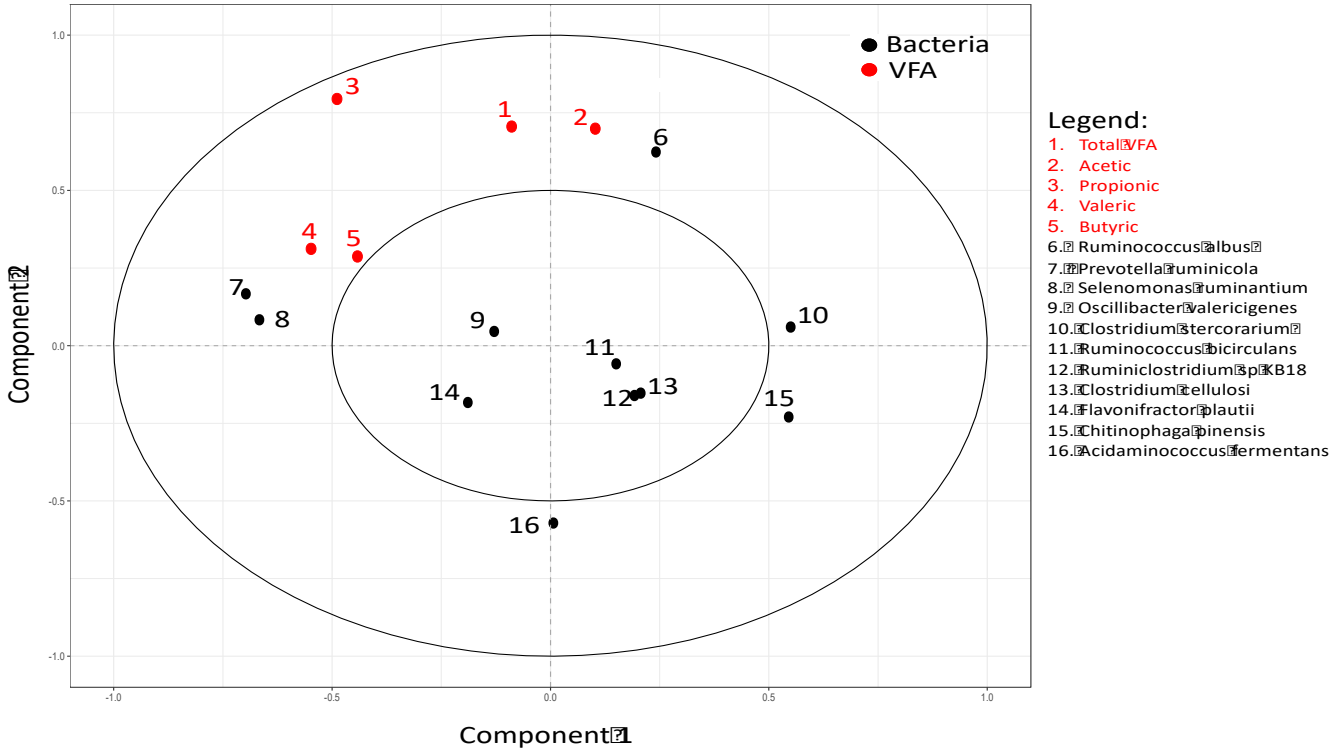

Bacteria
VFA
Component 2
Component 1
1.0
0.5
0.0
-0.5
-1.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Legend:
1. Total VFA
2. Acetic
3. Propionic
4. Valeric
5. Butyric
6. Ruminococcus albus
7. Prevotella ruminicola
8. Selenomonas ruminantium
9. Oscillibacter valericigenes
10. Clostridium stercorarium
11. Ruminococcus bicirculans
12. Ruminiclostridium sp KB18
13. Clostridium cellulosi
14. Flavonifractor plautii
15. Chitinophaga pinensis
16. Acidaminococcus fermentans

Figure 4.4. Heatmap and correlation circle plot generated from the output of regularized canonical correlation (rCC) method. (A) shows the correlations between VFAs (total VFA, acetic, butyric, propionic, and valeric acids) and rumen bacterial species in the first two rCC components. The color key indicates the correlation values among variables. (B) shows the correlation circle plot, where VFAs and bacteria are shown inside a circle of radius 1 centered at the origin, with strongly associated (or correlated) variables being projected in the same direction from the origin. Two circumferences of radius 1 and 0.5 were plotted to reveal the correlation structure of the variables.

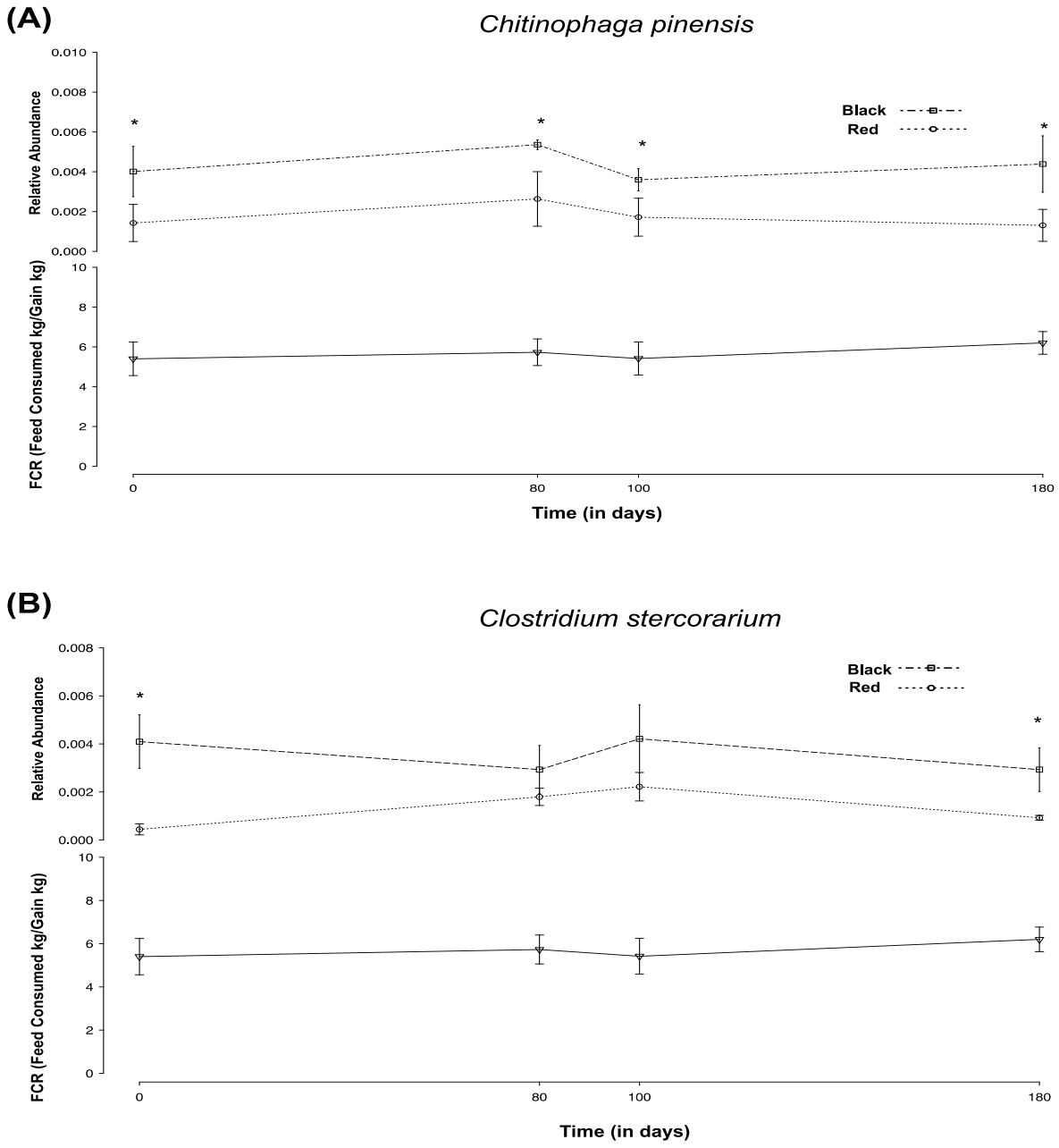


Figure 4.5 Differentially abundant bacterial species detected in Black and Red Angus when FCR (Feed consumed kg/Gain kg) was adjusted to time (in days). Unadjusted raw average relative abundance and standard errors of (A) *Chitinophaga pinensis*, (B) *Clostridium stercorarium*, and (C) *Ruminococcus albus* detected when FCR (kg/day) was adjusted to time (in days). Statistical procedures were performed by ANCOM (calculated on the log-ratio matrix; FDR, $q < 0.1$).

Chapter 5

Discovery and targeted functional profiling of novel glycoside hydrolases through selective pressure on the native rumen microbial community

5.1 Introduction

Comprised mainly of cellulose, hemicellulose, and lignin, lignocellulose is a major constituent of plant cell walls and is the most abundant organic polymer on Earth (McNeil et al. 1984). However, the suitability of lignocellulose as a substrate for the synthesis of biobased products is limited by its resilience to enzymatic digestion (Chafe 1969, McNeil et al. 1984). A specialized group of enzymes, known as carbohydrate-active enzymes (CAZymes), are required for the efficient degradation of lignocellulosic biomass in order to access the carbon sources within the lignocellulosic matrix (Cantarel et al. 2009, Lombard et al. 2014). CAZymes are a widespread and structurally diverse set of enzymes involved in the breakdown, biosynthesis or modification of lignocellulose and can be produced by microorganisms that inhabit various microbiomes including the mammalian gut (Lombard et al. 2014). The potent lignocellulolytic capability of CAZymes is conferred by an array of enzymatic catalysts, usually grouped into families according to amino acid sequence similarity (Lombard et al. 2014), which facilitate the degradation of complex polymers into simple sugars. The major CAZymes that degrade carbohydrate polymers are glycoside hydrolases - GHs, polysaccharide lyases - PL and polysaccharide monooxygenases (Munoz-Munoz et al. 2017). GHs are the most abundant CAZymes and are widely employed in biotechnological and biomedical settings (Cantarel et al. 2009). Yet, the discovery of novel GHs from microbial communities has been a challenging task due to the complexity and diversity of CAZyme families present in microbial habitats, which result in many as-yet-uncharacterized members in these families.

Microbial communities are dynamic and can evolve novel CAZymes according to the external environment and the nature of substrates available for metabolism (Wilkens

et al. 2017). This attribute enables us to use selective pressure to enrich the microbial environment with substrates that favor the growth of microbes that express the desired biocatalyst. One fascinating microbial community that has garnered much interest due to its phylogenetic diversity is the one residing in the forestomach of ungulates, termed rumen microbiome (Shaani et al. 2018, Russell and Rychlik 2001). This microbial ecosystem is an ideal model for the discovery of novel enzymes as it is contained within a unique animal host and can be easily manipulated through dietary interventions to gain insight into the dynamics and functions of the microbiota (Weimer 2015). Rapid advances in omics technologies has facilitated the discovery of thousands of genes encoding novel CAZymes in the rumen microbiome during the past two decades (Brulc et al. 2009, Hess et al. 2011, Wang et al. 2013, Svartström et al. 2017, Gharechahi and Salekdeh 2018, Meng Qi et al. 2011, Comtet-Marre et al. 2017, Fuyong Li and Le Luo Guan 2017). However, only a handful of these candidate enzymes have entered industrial application, mainly due to annotation mistakes (sequence/activity incoherence) in publicly databases (Fernández-Arrojo et al. 2010, Ferrer et al. 2016). This technical issue is further exacerbated by the fact that individual proteins within a protein family can have vastly diverse functions (Franzosa et al. 2015, Schnoes et al. 2009, Levin et al. 2017).

Targeted functional profiling of a microbial community (Kaminski et al. 2015) following metatranscriptomic sequencing may overcome these limitations as it allows the compilation of a *de novo* database of marker peptides derived from reference proteins of interest. This was recently demonstrated in the discovery of a novel biomarker of host-microbial symbiosis in the human gut (Levin et al. 2017). However, the complexity and diversity of CAZymes in microbial habitats has made targeted functional profiling of CAZymes extremely difficult, and consequently there remains substantial gaps in our knowledge of the functions of uncharacterized members of the CAZyme families.

Here, we adopted a selective pressure approach to enrich the rumen ecosystem of Angus bulls fed a forage-rich diet for microbes capable of degrading lignocellulose, in order to facilitate mining of the rumen metatranscriptome for metabolically active CAZyme families. Next, we employed the targeted functional profiling of these families to capture ecological information (abundance and distribution) of functionally distinct members of

enriched CAZyme families to guide the discovery of unknown enzymes. By screening 775 uncharacterized members of the GH11 family and 409 of the GH45 family, we identified 18 putative xylanases (GH11) and three putative endoglucanases (GH45). These outcomes highlight the usefulness of this strategy for uncovering novel rumen enzymes for the breakdown, biosynthesis or modification of lignocellulosic biomass.

5.2 Materials and methods

5.2.1 Animal trial and sampling

The experimental procedures described here were reported previously (Thompson 2015), with the study protocol being reviewed and approved by the University of Manitoba animal care committee. Briefly, rumen contents were collected from 12 purebred bulls (mean age of 249 ± 22 days and average body weight of 313.9 ± 32 kg) raised in confinement at the Glenlea Research Station (University of Manitoba) according to the guidelines of the Canadian Council on Animal Care (CCAC) (CCAC 1993). In the current trial, bulls were fed a forage based diet (Supplementary Table 1) throughout the experimental period (180 days). Then, representative samples (250 ml) of rumen contents (liquid and solid fractions) were collected over four-time points (0, 80, 100, 180 d) (48 samples in total) using a Geishauser oral probe (Duffield et al. 2004), immediately snap frozen in liquid nitrogen, and stored at -80°C for further processing. The feed intake of individual bulls was recorded using the GrowSafe® feeding system (GrowSafe Systems Ltd., Airdrie, Alberta, CA) to calculate feed conversion rate (FCR), which is a ratio of dry matter intake to average daily gain (computed on a biweekly basis (Montanholi et al. 2010)). The bulls were then ranked into two groups: high and low FCR, with high (H-FCR) and low (L-FCR) standing for inefficient and efficient cattle in terms of diet utilization, respectively.

5.2.2 RNA extraction and sequencing

Total RNA extraction and sequencing protocols were described by Fuyong Li and Le Luo Guan (2017). In summary, total RNA was extracted from rumen samples using the TRIzol protocol based on the acid guanidinium-phenol-chloroform method (Chomczynski and Sacchi 2006, Béra-Maillet et al. 2009) with modifications reported previously (Fuyong Li and Le Luo Guan 2017). Approximately 200 mg of rumen sample was subjected to RNA

extraction with the addition of 1.5 ml of TRIzol reagent (Invitrogen, Carlsbad, CA, USA), followed by 0.4 ml of chloroform, 0.3 ml of isopropanol, and 0.3 ml of high salt solution (1.2 M sodium acetate, 0.8 M NaCl) for the extraction protocol. Total RNA (100 ng) of each sample was used for library construction using the TruSeq RNA sample prep v2 LS kit (Illumina, San Diego, CA, USA) without the mRNA enrichment step. Finally, cDNA fragments (~140 bp) were paired-end (2 X 100 bp) sequenced using an Illumina HiSeq 2000 system.

5.2.3 Bioinformatic and statistical analysis

After checking the quality of fastq-formatted sequences using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), the software Trimmomatic (version 0.32; 58) was used to trim residual artificial sequences, cut bases with quality scores below 20, and remove reads shorter than 50 bp (Bolger et al. 2014). Then, SortMeRNA (Kopylova et al. 2012) (version 1.9) was used to sort the filtered reads into fragments of mRNA for microbial classification. Prior to performing the taxonomic classification, a Kraken (Wood and Salzberg 2014) standard database was built based on all complete genomes of bacteria downloaded from NCBI plus complete genomes from organisms isolated from the rumen or from ruminant feces or saliva deposited in the Hungate1000 project (JGI's IMG database) (Neves et al. 2017). Then, each pair of mRNA sequences was assembled by MEGAHIT (Li et al. 2015), with the resulting contigs being assigned by Kraken (through k-mer discrimination) to the lowest common ancestor in a customized standard database for microbial classification. For gene annotation, the contigs were submitted to MG-RAST (Meyer et al. 2008a) where they were de-replicated and quality checked using the methods described by Gomez-Alvarez *et al.* (Gomez-Alvarez et al. 2009) and Cox *et al.* (Cox et al. 2010). The per-base coverage depth across all contigs was calculated by mapping raw reads from each sample against the assembled contigs using BBMap v35.92 with the parameters 'kfilter=22, subfilter=15 and maxindel=80' (https://sourceforge.net/projects/bbmap/). Within MG-RAST, the contigs were annotated using subsystems technology (Aziz et al. 2008) with maximum e-value of 10^{-5}, minimum percent identity of 60, and minimum alignment length of 30 as the parameter settings (Xu et al. 2015). Translated non-rRNA sequences from the mRNA-

enriched RNA sample were submitted to a local version of dbCAN (Yin et al. 2012) to annotate sequences for the presence of CAZymes with a cutoff e-value < 10^{-5}. ShortBRED was then used to determine the abundance of distinct members of CAZymes of interest in the metatranscriptomic dataset, with the sequences being grouped at a specified amino acid similarity threshold of 85% identity to detect non-redundant representative matches (Kaminski et al. 2015). UniRef90 (downloaded on March 2017) was used as the comprehensive protein reference catalog to annotate the representative short peptides (markers) within the CAZyme family (Suzek et al. 2015). The theoretical atomic models of the markers quantified in the previous step were constructed using I-TASSER (Zhang 2008, Yang et al. 2014). Multiple threading alignments of the markers were generated by the meta-server LOMETS (Wu and Zhang 2007) to identify template structures from the Protein Data Bank (PDB) library, followed by structural assembly and refinement steps, with subsequent reconstruction of the atomic models (Li and Zhang 2009). Protein-ligand binding sites of the homology models were verified with the I-TASSER associated COACH package (Yang et al. 2013). Read counts classified by kraken (microbial taxonomic assignment), MG-RAST (functions) and dbCAN (CAZyme families) were subjected to differential abundance analysis (L-FCR vs. H-FCR) using edgeR (functions: *glmFit* to fit the negative binomial generalized log-linear models and *glmLRT* to conduct likelihood ratio tests for coefficient contrasts) (Robinson et al. 2010, Robinson and Smyth 2008, McCarthy et al. 2012). The trimmed mean of M-values (TMM) method was used to normalize the data and minimize the log-fold change between samples (Bullard et al. 2010). All *P* values were corrected for a false discovery rate (FDR) of 0.05 using the Benjamin-Hochberg algorithm (Benjamini and Hochberg 1995), and FDR-corrected *P* values <0.05 were considered as significant. Cladograms were generated by GraPhlAn (Asnicar et al. 2015) and the heat trees were built using the package *metacodeR* (Foster et al. 2017). All statistical procedures were done in R 3.4.2 (R Core Team, 2017) and Python 3.6.0.

5.2.4 Cloning, protein expression and purification

The sequence of the gene of putative xylanase 1 selected from candidates obtained in the previous bioinformatic analysis was cloned in *pET43.1a* (ligation at *Xho1* and *BamH1*

sites) using commercial services (Genscript, NJ, USA). The protein expression vector was then transformed into chemically competent *E. coli* Rosetta-gami™ 2 DE3 cells (Millipore, Ontario, Canada), with single colonies being grown in Luria broth medium supplemented with 100 µg/µL of ampicillin (Amresco, Solon, OH) at 37°C until reaching OD_{600} of 0.5-0.6. Overexpression was induced with 0.4 mM IPTG, and the cultures were grown for 8 h at 24°C. Cells were then harvested by centrifugation at 6000 × g for 15 min at 4°C. The cell pellets were resuspended in phosphate-buffer (pH 7.4), containing 1mM PMSF (phenylmethylsulfonyl fluoride) and lysed using Emulsiflex (Avestin, Ottawa, Canada) at a pressure of 206.8 MPa. The unbroken cells and cell debris were pelleted by centrifugation at 17,000 × g for 30 min. The supernatant was then incubated with subtilisin resin (Profinity exact Expression Technology, Bio-RAD, USA) for 1 h at 4ºC, and unbound proteins were washed away with phosphate buffer. The protein of interest was eluted by incubating the resin overnight with elution buffer (pH 7.2, 0.1M Sodium phosphate and 0.1M Sodium fluoride) at 4°C, and then dialyzed (Spectra/Por membrane tubing, Vol/Length: 1 ml/cm) against McDougall's buffer, pH 7.0 (McDougall 1948) and concentrated using 10,000 MWCO concentrators (Millipore, USA) to 0.4-1 mg/ml. The concentrated protein samples were aliquoted, flash-frozen and stored at −80°C. Protein concentration was determined by colorimetric detection and quantification of total protein using the Pierce BCA protein assay kit (Thermo-Fisher Scientific) with the bovine serum albumin as standard. The purified protein was then visualized by SDS-PAGE gel.

5.2.5 Differential scanning fluorimetry (DSF)

To investigate the effect of pH on protein stability, DFS assay of Xylanase 1 in different buffers (100 mM Sodium acetate buffer: pH 5.0, 6.0; 100 mM Tricine buffer: pH 7.0, 8.0, and 9.0; McDougall's buffer: pH 6.0, 7.0, and 8.0) spanning the interval of pH 3–8 was performed. Xylanase 1 in a final concentration of 5 µM was mixed with SyproOrange dye (Thermo Fisher Scientific, USA). Prior to use, the dye stock was diluted 1:50 (100X) in water and used immediately while protecting from light to reduce photobleaching. The optimal dilution of dye in the assay was determined empirically with a 5X dilution for the final assay. The thermal denaturation assay was performed in a total volume of 40 µl. All samples were run in duplicates. The thermal scan was conducted from 25 to 95°C, at

0.5 °C/min (ViiA 7 Real-Time PCR System, ThermoFisher). The melting point (T_m) was calculated by fitting the raw fluorescence data over the temperature using the Boltzmann equation in GraphPad Prism program (GraphPadPrizm 7 for Windows, GraphPad Software, USA).

5.2.6 Size-exclusion chromatography

The oligomeric state and homogeneity of Xylanase 1 was determined by size-exclusion chromatography on Superdex 75 (10/30) column (GE Healthcare, Canada), equilibrated with McDougall's buffer, pH 6. Molar mass of the protein peak was calculated using a logarithmic interpolation of elution volumes (Ve) using a gel filtration LMW calibration kit (GE Healthcare, Pittsburgh,USA) containing 1) blue dextran 2000 (V_0), 2) thyroglobulin (670 kDa), 3) g-globulin (158 kDa), 4) ovalbumin (44 kDa), 5) myoglobulin (17 kDa), and 6) vitamin B12 (1.3 kDa).

5.2.7 Kinetic measurements

Xylanase 1 activity was determined by measuring the quantity of reducing sugars (xylose, molecular weight: 150 g/mol) released from xylan (Beachwood xylan, Megazyme) by the dinitrosalicylic acid (DNS) method (Miller 1959). Before kinetic calculations, all the parameters for the assay were optimized. The minimal concentration of the enzyme that produced a linear dependence of generated product with the time was chosen, as well as the minimal time of reaction within the linear part of the curve. For kinetic measurements Xylan was incubated at 40°C with activity buffer - McDougall's buffer (McDougall 1948), pH 6) for 10 min for equilibration and then the purified Xylanase was added and the reaction was performed for 10min. The final concentration of enzyme was fixed at 0.05 µM, and the final concentration of Xylan varied (0, 0.88, 1.75, 3.5, 7.0, 15.0, and 30.0 mg/ml). The total volume of reaction was 200 µl. The samples with the same concentrations of substrate but without enzyme addition were treated the same way and were used as negative controls. After adding 300 µl of DNS reagent to stop the reaction, the samples were boiled for 5 minutes, and then put on ice, following the measurement under absorbance at 540 nm using a plate reader (SpectraMax M3). All reactions were performed in duplicates. The plots of the reaction velocity against the

corresponding substrate concentration were fitted with Michaelis–Menten equation ($v_0 = k_{cat}[E]_0[S]_0/([S]_0 + K_M)$) using GraphPad Prism program (GraphPadPrizm 7 for Windows, GraphPad Software, USA).

5.2.8 Thermal inactivation of Xalanase 1 and thermodynamic analysis

Thermal inactivation assay was performed at 25, 40, 50 and 60^{o}C. The samples, containing 0.5µM of Xalanase1 in McDougall's buffer (pH 6.0) were incubated at the specified temperatures. Then, 20µl aliquot was taken out at each time point and stored on ice until the activity measurements were performed as described above, using 0.05µM of Xalanase 1 and 30 mg/ml of xylan. A non-heated enzyme was used as positive control and its activity was taken as 100%.

Enzyme inactivation over time was described as a first-order reaction:

$$ln\ A/A_0 = -kt \quad (1)$$

where A – activity at time t, A_0 – initial activity at time zero, k is inactivation rate constant at the tested temperature (min^{-1}) and t is time (min). k values were calculated from linear regression analysis of the natural logarithm of residual activity versus incubation time and replotted in Arrhenius plot. Activation energy (Ea) was calculated using the slope of Arrhenius plot according to Eq. 2

$$ln(k) = -Ea/RT+c \quad (2)$$

where R is the gas constant (8.314 J mol^{-1} K^{-1}) and T is the absolute temperature.

The half-life of Xalanase 1 ($t_{1/2}$ in min), defined as time after which activity is reduced to one-half of its initial value was determined according to

$$t_{1/2} = ln\ (2)/k \quad (3)$$

The D-value is the time (min) needed to reduce the initial activity to 90%. It is inversely related to k-values and mathematically expressed as

$$D = ln(10)/k \quad (4)$$

The values of Gibbs free energy (ΔG^o, kJ mol^{-1}), enthalpy (ΔH^o, kJ mol^{-1}), and entropy ΔS^o (J $mol^{-1}K^{-1}$) were determined as

$$\Delta G^o = -RTln(kh/k_bT) \quad (5)$$

$$\Delta H^o = Ea-RT \quad (6)$$

$$\Delta S^o = (\Delta H^o - \Delta G^o)/T \quad (7)$$

Where h is the Plank constant (6.626 x 10^{-34} Js) and k_b is the Boltzmann constant (1.38 x 10^{-23} JK^{-1}). All experiments were performed in duplicates.

5.3 Results

5.3.1 Taxonomic composition of the rumen microbiota revealed by metatranscriptomic sequencing

Assembly of mRNA reads resulting from total RNA sequencing of rumen contents yielded on average 7,627 contigs per sample (with an average extension of 474.3 ± 26.67 bp and N50 of 462.6 ± 27.99 bp), which were further classified using a taxonomic assignment strategy as described in chapter 3 (Neves et al. 2017). Approximately 51% of the quality filtered mRNA reads were mapped to the assembled contigs, indicating that they represented a significant proportion of the metatranscriptome (Supplementary Data 1). Of the 20 phyla identified in the dataset, the majority of sequences were assigned to Bacteroidetes (45%), followed by Firmicutes (23%), Proteobacteria (14%), Spirochaetes (5.0%), Verrucomicrobia (2.3%), Actinobacteria (2.2%), Tenericutes (2.1%), and Fibrobacteres (1.4%) (Figure 5.1a; Supplementary Data 2; and Supplementary Figure 1).

5.3.2 Functional capability of the rumen microbiota

To examine the functional potential of the microbial community associated with the degradation of lignocellulose, mRNA transcripts (assembled reads) were mapped against the publicly available Subsystems database using MG-RAST (Meyer et al. 2008b), which resulted in the detection of 1205 functions (ranging from the most detailed, level 4, to the least detailed category, level 1) (Supplementary Data 3). Central carbohydrate metabolism (including glycolysis/gluconeogenesis, glyoxylate cycle, pyruvate metabolism and pentose phosphate pathway) and protein biosynthesis were the most abundant functional categories, representing 10% and 33% of the annotated reads, respectively

(Supplementary Figure 2). In the polysaccharides- and monosaccharides-related categories (level 3), the cellulosome and xylose utilization systems comprised 0.39% and 0.72% of the total annotated reads, respectively (Figure 5.1b).

Assembled metatranscriptomic contigs were also aligned against the CAZyme database (Yin et al. 2012) to obtain more in-depth information regarding the carbohydrate enzymes in the dataset. A total of 6904 unique alignments were recovered, dominated by GHs (42.8% of the total CAZyme matches) and carbohydrate-binding modules - CBMs (33.2%), followed by glycosyltransferases- GTs (9.5%), carbohydrate esterases - CEs (6.5%), Dockerin (2.7%), S-layer homology domains - SLHs (2.6%), PL (2.0%), Cohesin (0.5%), and auxiliary activities – AA6 (0.2%) (Figure 5.1c).

The rumen metatranscriptome of forage-fed bulls was notably enriched with GH catalytic modules that provided 2914 hits belonging to 61 families (Supplementary Data 4). Cellulases (GH5, GH9, GH45, and GH48) and hemicellulases (GH8, GH10, GH11, GH26, GH28, GH53) exhibited a high representation (25% of the total CAZyme matches) in the metatranscriptome, indicating that they were actively involved in the degradation of plant cell wall components contained in the forage diets. A wide variety of non-catalytic proteins known as CBMs were highly represented (2297 hits) and were predicted to partake in interactions with various substrates such as cellulose (e.g., CBM1, CBM2, CBM3, CBM6, CBM13, CBM16, CBM44), xylan (e.g., CBM4, CBM22, CBM37), starch (e.g., CBM20, CBM26), and chitin (e.g., CBM50) (Supplementary Data 4). Other important classes of CAZymes frequently encountered in our dataset were CEs (e.g., CE1, CE2, CE3, CE4, CE7, CE12) and PLs (e.g., PL1, PL9, PL11). The presence of accessory modules (>400 hits in this study) commonly found in bacterial cellulosome-associated structures (AAs, dockerins, and cohesins) and the SLH domain, provided additional evidence of active cellulosome-mediated plant cell-wall degradation employed by rumen microorganisms (Supplementary Data 4).

5.3.3 Fibrolytic bacteria and glycoside hydrolases are abundant in the rumen of feed efficient animals

To facilitate the mining of active CAZyme families in the rumen metatranscriptome, we compared the rumen microbial population in feed efficient cattle with those that were less

effective at digesting the same diet. To achieve this, we divided our experimental bulls into two groups based on feed efficiency ranking (measured by feed conversion ratio, FCR): 1) efficient or low feed conversion rate (L-FCR) and 2) inefficient or high-FCR (H-FCR). The FCR of the two groups of animals was statistically divergent ($P < 0.01$), with L-FCR bulls exhibiting a feed conversion efficiency 22% lower than H-FCR bulls (Supplementary Figure 3).

Although there was no difference in the bacterial diversity between the feed efficiency groups (Supplementary Figure 4), the data showed that fibrolytic bacteria and GHs were abundant in the rumen of feed efficient animals. Of the 115 species analyzed in all samples (Supplementary Data 2), *Fibrobacter succinogens* (a cellulose degrader species (Suen et al. 2011)) was the only species affected by feed efficiency, showing a nearly 0.5-$\log_2$-fold change increase ($P < 0.05$) in L-FCR relative to H-FCR (Figure 5.2a). Our results also showed that the relative abundance of *Fibrobacter succinogens* and other common plant cell wall degraders (*Butyrivibrio proteoclasticus* and *Ruminiclostridium* sp KB18) exhibited > 4-$\log_2$-fold change increase on the 180th day relative to day 0 in L-FCR compared to H-FCR (Figure 2b). *Prevotella ruminicola* exhibited a 0.6-$\log_2$-fold change increase on the 180th day relative to day 0 in L-FCR compared to H-FCR (Figure 5.2b).

Next, we examined whether the functional potential of the microbial community could be linked to the feed efficiency groups, with the aim of finding genes involved in the degradation of lignocellulosic biomass (Figures 5.1d-e). Of the 1205 features analyzed in all samples (Supplementary Data 3), we could not confirm genes for lignocellulosic biomass degradation differentiating the rumen microbiome of L-FCR from H-FCR. Instead, we found ammonia assimilation functions mediated by aspartate-ammonia ligases (EC: 6.3.1.1) (Ricard et al. 2006) and motor organelles, which propel the rotating flagella to enable bacteria to move towards favorable environments (Sowa and Berry 2008), at a higher abundance in the rumen of L-FCR compared to H-FCR (Figure 5.3d). However, genes involved in degradation of di- and oligosaccharides, which includes cellulose (e.g., Cellobiose phosphorylase - EC: 2.4.1.20) and xylose utilization (e.g., Endo-1,4-β-

xylanase), exhibited a >5-log_2-fold change increase in their abundance on the 180th day relative to day 0 in L-FCR compared to H-FCR (Figure 5.2e; Supplementary Data 5).

To further investigate whether feed efficiency affected CAZymes, we then analyzed all CAZymes families in our dataset in relation to the FCR groups (Figure 5.2c; Supplementary Data 4). As observed in the results of the functional profiles, we did not detect CAZymes differences between L-FCR and H-FCR; however, we found a consistent 1.9-log_2-fold change increase in the relative abundance of GH13 on the 180th day relative to day 0 in L-FCR compared to H-FCR (Figure 5.2c). GH13, the most abundant CAZyme family in the present study with 6.8% of the total matches, is a well-represented CAZyme family in the rumen (Hess et al. 2011) and the largest glycoside hydrolase family, acting mainly on the catalysis of α-glucoside linkages encapsulated in starch and glycogen. More importantly, we found that GH11 (endo-β-1,4-xylanase - EC 3.2.1.8), GH45 (endoglucanase: EC 3.2.1.4), and CBMs connected to cellulose degradation (CBM79) exhibited >2.5-log_2-fold change increase in their relative abundance in L-FCR compared to H-FCR on day 180 relative to day 0 (Figure 5.2c).

5.3.4 Targeted functional profiling of GH11 and GH45 families

To gain insight into functionally distinct members of GH11 (xylanases) and GH45 (endoglucanases), we applied ShortBRED (Kaminski et al. 2015) to screen those families against a *de novo* protein reference database built from UniProt, and then profile their abundance and distribution in the rumen metatranscriptome. By screening 775 uncharacterized members of the family GH11 and 409 of the family GH45, we identified 18 putative xylanases (GH11) and three putative endoglucanases (GH45) (Supplementary Data 6).

In this study, bacteria and eukaryotic organisms represented the major sources of the identified xylanases and endoglucanases (Figure 5.3b). While only two genes were sourced by known rumen microbes such as *Fibrobacter succinogens* (100% identity; UniProt IDs: C9RR38 and D9SBI1) and uncultured rumen ciliates (UniProt ID: G5DDC1; 70.5% identity with *Epidinium caudatum,* Supplementary Figure 5), the vast majority of enzymes matched bacteria and fungi strains found in other environments (e.g., soil), indicating that they are unknown in the rumen microbiome (Figure 3b).

5.3.5 Structural analysis of Xylanase 1 through homology modelling

To further investigate key active sites and the tertiary conformation of xylanase 1 (the most broadly distributed and abundant xylanase in our rumen metatranscriptome dataset), we constructed and compared a homology model of xylanase 1 with crystal structures deposited on the Protein Data Bank (PDB) using I-TASSER (Zhang 2008, Yang et al. 2014). Of the 33 crystal structures reported for the family GH11 in PDB and CAZyme databases, the top homology model of xylanase 1 showed 57% sequence identity with the crystal structure of a xylanase family 11 (PDB ID: *1h4hB*) encoded by *Bacillus agaradhaerens* (Supplementary Figure 6; Supplementary Table 2). The confidence score (C-scores) of the top model for xylanase 1 was 1.67, indicating good quality of the predicted homology model (C-score is typically in the range [-5, 2], where scores of higher values signify a model with high confidence). Superimposing the homology model of xylanase 1 onto the crystal structure of *Bacillus agaradhaerens* xylanase resulted in global structural alignment scores (TM-scores) of 0.95 (TM-score >0.5 indicates a model of correct topology) and root-mean-square deviation (RMSD) of the TM-aligned residues of 0.47 Å (Supplementary Figure 6; Supplementary Table 2), confirming that the model was in agreement with the crystal structure of *Bacillus agaradhaerens* xylanase.

Xylanase 1 exhibited the β jelly-roll fold typical of GH xylanases, with the concave antiparallel β sheet being constituted of nine β strands (β2, β3, β6, β8, β9, β10, β11, β12, β14) and the antiparallel convex sheet comprising six β strands (β1, β4, β5, β7, β13, β15). The α – helixes were found in the loops connecting β strands 6 and 7 and β strands 13 and 14 (Figure 5.3c). The structure of xylanase 1 resembles the shape of a right hand (Törrönen et al. 1994) with the “fingers” at the top of the palm, comprising the loops connecting β1 to β2, β3 to β4, β5 to β6, β14 to β15 and β7 to β8 and the “thumb”, consisting of the loop connecting β11 to β12, at the right-hand side of the molecule (Vardakou et al. 2008). Multiple functional annotations of xylanase 1 was investigated with the I-TASSER associated COACH package (Yang et al. 2013), which showed the ligand 1,2-Deoxy-2-Fluoro-Xylopyranose (DFX) (PDB ID: *1c5iA*) (Joshi et al. 2000) docked in the predicted substrate-binding cleft (Figure 5.3d).

Following the procedures described above, we also constructed homology models (TM scores ≥ 0.7) for other xylanases and endoglucanases found in the rumen metatranscriptome (Supplementary Figures 6 and 7; Supplementary Table 2), which provides a basis for understanding the molecular functions of these enzymes and guide future structure determination.

5.3.6 Purification and characterization of oligomeric state of Xylanase 1

The purification resulted in a highly pure enzyme with an expected molecular weight of 23 kDa (Figure 5.4a). Gel-filtration analysis revealed that Xylanase 1 was eluted as a single homogeneous peak with a calculated molecular weight of 25 kDa, suggesting that Xylanase 1 exists in a monomeric state in solution (Figure 5.4b).

5.3.7 Kinetic analysis of Xylanase 1

The pH curve exhibited a standard bell-shaped curve with an optimum pH of 6.0, with the enzymatic activity being still high at pH 7.0 and decreasing only at pH 8.0 (Figure 5.4c). The catalytic parameters for the activity of Xylanase 1 towards xylan determined by the Michaelis-Menten equation were as follows: a) cleavage rate and catalytic efficiency of 480s^{-1} and 872 $M^{-1}s^{-1}$, respectively, and b) *Km* of 8.7 ± 0.9 mg/ml (Figure 5.4d).

5.3.8 Thermal stability of Xylanase 1

To assess favorable conditions for the thermal stability of Xylanase 1, a range of pHs and different buffers were screened. As observed in Figure 5D, the melting temperatures (*Tm* values) at pHs 6.0 and 7.0 in all buffer systems provided the most stable environment for Xylanase 1 (Figure 5.5a). Additionally, the residual activity of Xylanase 1 was evaluated at 25, 40, 50, and 60ºC by thermal inactivation assays in a time- and temperature-dependent manner (Table 1; Figures 5.5 b-c). The semi-log plots of the residual activity versus heating time were linear at all temperatures studied (Figures 5.5 b-c), demonstrating that inactivation of Xylanase 1 is a simple first-order monophasic process.

Inactivation rate constants (*Kd*) for each temperature were obtained from the slopes of the Arrhenius equation (Figure 5.5c), with *Kd* values increasing ~10-fold per 10ºC, suggesting a high degree of irreversible denaturation (Table 1). Additionally, the

results showed that Xylanase 1 was stable at 25 °C and had activity loss of only 10% after 1 hour of incubation (*D* value of 1354 min) (Table 1). At 40 °C, 35% of activity remained after 1 hour of incubation, and the *D* value was of 127 min. The loss of activity with a *D* value of 2.5 min was observed after incubating the enzyme at 60 °C (Table 1).

The thermodynamic parameters of inactivation including the Gibbs free energy change (*ΔG*), the enthalpy change (*ΔH*), and the entropy change (*ΔS*) were also assessed to understand the enzyme's behavior at each step of the heat-induced denaturation process (Table 1). The value of *ΔH* was 150.52 kJ mol-1 (25 °C), and this result was independent of temperature, assuming that there is no change in the enzyme heating capacity. The fact that *ΔH* values were positive indicates that enzyme inactivation is an endothermic process. The results also showed that *ΔG* value reduced significantly from 78.64 to 70.75 kJ mol-1 when the incubation temperature increased from 25 to 60 °C, showing that protein destabilization followed the rise in the temperature. Moreover, the values of *ΔS* were positive meaning that there is an increase in the molecule disorder during the exposure to higher temperatures and that unfolding is a rate-limiting step for thermal inactivation (Table 1).

5.4 Discussion

The rumen microbiome has proven to be a valuable reservoir of microbial proteins with application in the biotechnological industry, most notably those involved in the hydrolysis of lignocellulosic biomass (CAZymes) (Seshadri et al. 2018). Identifying novel candidate proteins from the rumen microbiome (and other host sites) has typically been performed via comparison to sequence data, but this does not capture the nuanced differences in the functions of individual proteins within each family of CAZymes, due to vast structural diversity among members (e.g., GHs). In addition, such an approach to identify novel enzymes is also time-consuming. To circumvent these limitations, the present study employed and adapted existing bioinformatic tools to detect novel glycoside hydrolase enzyme sequences in the rumen metatranscriptome of Angus bulls fed a forage diet. Targeted functional profiling (Kaminski et al. 2015) using predictive computational analyses was used to discover a broadly distributed novel ruminal GH, xylanase 1. Application of this approach on a larger scale may be a critical tool for further discovery

of novel CAZymes within the rumen and other host-associated microbiomes, which may have critical functions in feed digestion and host health, as well as applications in the biotechnology industry.

While the targeted functional profiling (Kaminski et al. 2015) has been used to examine uncharacterized enzymes found in the human microbiome (Levin et al. 2017), these studies have not profiled CAZymes in microbial communities. Similarly, although metagenomic and metatranscriptomic screening have been widely applied to discover CAZymes in various microbiomes (Ferrer et al. 2005, Ferrer et al. 2016, Hess et al. 2011, Comtet-Marre et al. 2017), to our knowledge the combination of selective pressure on the native microbial community with the targeted functional profiling has not been applied to quantitative metatranscriptomic analysis in order to accelerate the discovery of new enzymes. To facilitate the identification of metabolically active CAZymes in the rumen metatranscriptome, we adopted a selective pressure approach to enrich the rumen ecosystem of bulls fed a forage diet for microbes capable of degrading lignocellulose. The selective advantage to a microbe in the ruminal ecosystem arises in one of two ways: 1) increased dietary abundance of a specific substrate (selective pressure), favoured by the microbe or 2) niche specialization by a microbe (the lack of competition for the substrate of choice) (Weimer 1998). In the present study, the bulls were fed with a forage (and thus lignocellulose-rich) diet, and their rumens were dominated by microbes (e.g., Bacteroidetes, Firmicutes, Spirochaetes, and Fibrobacteres) and CAZymes (e.g., GH5, GH9, GH45, GH11) known to be responsible for digesting complex and recalcitrant plant polymers. These findings reflect those of previous studies investigating the rumen microbiome of forage-fed animals (Comtet-Marre et al. 2017, Dai et al. 2015, Gharechahi and Salekdeh 2018, Svartström et al. 2017), and confirm that offering animals a forage-rich diet selectively enriches the rumen microbiome for microbes and enzymes involved in plant cell wall hydrolysis. As well as the GH families listed above, we also noted the abundance of the SLH domain in our data, which is to the best of our knowledge the first time its existence has been documented in the rumen metatranscriptome of cattle, having been previously characterized in the camel rumen (Gharechahi and Salekdeh 2018). The SLH domain is part of a large multi-enzyme complex known as cellulosome (Artzi et al. 2016, Bayer et al. 1998)**,** and its presence in the bovine rumen lends further credence to

the hypothesis that feeding animals high-forage diets over a prolonged period of time selectively enriches the rumen with specialized fiber-degrading enzymes.

In comparing the rumen microbiome among cattle divergent for feed efficiency, we found that cellulose- and hemicellulose-degrading species (*Fibrobacter succinogenes, Butyrivibrio proteoclasticus* and *Ruminiclostridium* sp KB18) and specific CAZymes (GH13, GH11, GH45, CBM79) were more abundant in the metatranscriptome of efficient cattle. This suggests that the rumen microbiome of efficient cattle can harbor increased numbers of fibrolytic species and enzymes when undergoing a long-term forage feeding. This is an important finding and provides more robust evidence of a causal relationship between the rumen microbes and feed efficiency than has previously been reported (Fuyong Li and Le Luo Guan 2017).

Having characterized the presence of enzyme families which we expected to be enriched in the forage-fed rumen, we then applied this strategy to identify enzymes of unknown functions in the GH11 and GH45 families, which were one of the most enriched CAZyme families in the rumen microbiome of the feed efficient cattle. Although the activities of certain GHs have been investigated in the rumen (Vardakou et al. 2008, Jones et al. 2017), a large number of GH11 and GH45 members are as yet uncharacterized, and are typically excluded in analyses of the rumen proteome. GH11, unlike other xylanase families (e.g., GH10), represents only endo-β-1,4-xylanases whose function is to cleave β-1,4-xylosidic bonds between xylose monomers, whereas endo-1,4-β-glucanases of the family GH45 play a role in the hydrolysis of the 1,4-β-D-glucan chain. Although these enzymes act on lignocellulosic substrates (xylose and cellulose), they exhibited different abundances and distributions in the rumen, suggesting that they may perform distinct activities within the GH11 and GH45 families (Figure 5.3a). Taking this ecological context into consideration, 18 previously uncharacterized xylanases (GH11) and three uncharacterized endoglucanases (Gh45) were found in the rumen, especially Xylanase 1 that was the most broadly distributed, widespread and abundant enzyme (Figure 5.3a). Crucially, Xylanase 1 was detected in the metatranscriptome of all animals, indicating that it has a core function or functions within the rumen. Kinetic analysis and thermodynamic experiments confirmed that the abundant, uncharacterized GH11 is an

active and stable enzyme capable of degrading xylan, which is the most common hemicellulose and considered the second most abundant biopolymer in the plant kingdom (Ebringerová and Heinze 2000, Stephen 1983).

The homology model of xylanase 1 revealed a similar structural fold and catalytic residues as reported for *1h4hB,* but it showed different rearrangements in the loops that sculpture the active site, which can impact substrate binding and consequently enzyme activity and function. Further investigations showed that Xylanase 1 had two predicted catalytic residues, E92 (catalytic nucleophile) and E182 (catalytic acid-base), located on β strands 9 and 14 (Figures 5.3c-d). The location of these residues adopts a similar conformation in GH11 xylanases and is entirely consistent with the catalytic apparatus of a retaining glycoside hydrolase, which hydrolyzes glycosidic bonds by a double displacement mechanism (Davies and Henrissat 1995). In Xylanase 1, the residue in close spatial proximity to E182 (the catalytic acid-base) is N180, which characterize enzymes that display non-acidic pH optimum (Joshi et al. 2000). Our experiments confirmed that Xylanase 1 possessed a pH optimum similar to the rumen pH of forage-fed cattle (~pH 6.0) (Holtshausen et al. 2013), and thus it is likely that this enzyme plays a key role in the digestion of xylan in the rumen.

The discovery of new enzymes to break down lignocellulose into simple sugars is urgently needed, as market demands for enzymes with applications in industrial processes and animal feed are quickly increasing as a result of the cumulative impacts of feedstocks on the environment. Unfortunately, a considerable amount of agricultural residuals are underutilized due to the lack of an effective enzymatic system to degrade lignocellulose and release its constituent sugars for fermentation. The current study highlights the usefulness of combining selective pressure on the native rumen microbial community with the targeted functional profiling performed by well-established algorithms (Kaminski et al. 2015, Neves et al. 2017, Wood and Salzberg 2014, Meyer et al. 2008a, Yin et al. 2012). This approach revealed not only the diversity of bacteria and genes associated with efficient plant cell wall digestion but also facilitated the characterization of novel CAZyme family members, which may be critical in feed degradation. Applying this strategy and its underlying concepts in microbial ecology, nutrition, and bioinformatics,

we discovered previously uncharacterized hemicellulases and endoglucanases. The demonstration of the xylanolytic capacity of the most abundant and conserved member of the GH11 family (xylanase 1) present in the rumen metatranscriptome validates the power of this strategy in discovering lignocellulolytic enzymes from microbial environments. However, like any other enzyme, the structural basis of these new enzymes must be investigated in more details to consolidate their status as suitable candidates for the industrial enzyme market.

5.5 Conclusions

In summary, this study has demonstrated a robust pipeline for the discovery and characterization of novel CAZymes in the rumen microbiome. Furthermore, we provide evidence that the host feed efficiency phenotype may be mediated though evolution of the rumen microbiome in response to dietary pressure. This proof-of-concept approach may have many applications outside of animal agriculture, particularly in the recovery of novel microbial enzymes for use in the biotechnology sector. It may be adapted to any microbial environment for the discovery of CAZymes of interest, provided that the targeted microbiome is easy to manipulate and facilitates enrichment for the microbes of interest. It is likely that in the future, many more lignocellulolytic enzymes with high efficiency will be discovered through this strategy as knowledge of the precise factors which drive microbial community shifts increases.

5.6 Literature cited

Artzi, L., Bayer, E.A., and Moraïs, S. (2016). Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nature Reviews Microbiology* 15**,** 83. doi: 10.1038/nrmicro.2016.164.

Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *Peerj* 3. doi: 10.7717/peerj.1029.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., et al. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9(1)**,** 75. doi: 10.1186/1471-2164-9-75.

Bayer, E.A., Chanzy, H., Lamed, R., and Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology* 8(5)**,** 548-557. doi: https://doi.org/10.1016/S0959-440X(98)80143-7.

Benjamini, Y., and Hochberg, Y. (1995). CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. *Journal of the Royal Statistical Society Series B-Methodological* 57(1), 289-300.

Béra-Maillet, C., Mosoni, P., Kwasiborski, A., Suau, F., Ribot, Y., and Forano, E. (2009). Development of a RT-qPCR method for the quantification of Fibrobacter succinogenes S85 glycoside hydrolase transcripts in the rumen content of gnotobiotic and conventional sheep. *Journal Of Microbiological Methods* 77(1), 8-16. doi: 10.1016/j.mimet.2008.11.009.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 30(15), 2114-2120. doi: 10.1093/bioinformatics/btu170.

Brulc, J.M., Antonopoulos, D.A., Berg Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., et al. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences* 106(6), 1948-1953. doi: 10.1073/pnas.0806191105.

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11(1), 94. doi: 10.1186/1471-2105-11-94.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* 37, D233-D238. doi: 10.1093/nar/gkn663.

CCAC (1993). "Guide to the care and use of experimental steers," in *Canadian Council on Animal Care*. 2nd ed. ed (Ottawa: Ottawa, Ontario, Canada).

Chafe, S.C. (1969). The fine structure of the collenchyma cell wall. *Planta* 90(1), 12-21. doi: 10.1007/BF00389292.

Chomczynski, P., and Sacchi, N. (2006). The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nature Protocols* (2), 581.

Clark, J.H., Budarin, V., Deswarte, F.E.I., Hardy, J.J.E., Kerton, F.M., Hunt, A.J., et al. (2006). Green chemistry and the biorefinery: a partnership for a sustainable future. *Green Chemistry* 8(10), 853-860. doi: 10.1039/B604483M.

Comtet-Marre, S., Parisot, N., Lepercq, P., Chaucheyras-Durand, F., Mosoni, P., Peyretaillade, E., et al. (2017). Metatranscriptomics Reveals the Active Bacterial and Eukaryotic Fibrolytic Communities in the Rumen of Dairy Cow Fed a Mixed Diet. *Frontiers in Microbiology* 8(67). doi: 10.3389/fmicb.2017.00067.

Cox, M.P., Peterson, D.A., and Biggs, P.J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11(1), 485. doi: 10.1186/1471-2105-11-485.

Dai, X., Tian, Y., Li, J., Su, X., Wang, X., Zhao, S., et al. (2015). Metatranscriptomic Analyses of Plant Cell Wall Polysaccharide Degradation by Microorganisms in the Cow Rumen. *Applied and Environmental Microbiology* 81(4), 1375.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* 3(9), 853-859. doi: https://doi.org/10.1016/S0969-2126(01)00220-9.

Duffield, T., Plaizier, J.C., Fairfield, A., Bagg, R., Vessie, G., Dick, P., et al. (2004). Comparison of techniques for measurement of rumen pH in lactating dairy cows. *Journal of Dairy Science* 87(1)**,** 59-66.

Ebringerová, A., and Heinze, T. (2000). Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties. *Macromolecular Rapid Communications* 21(9)**,** 542-556. doi: doi:10.1002/1521-3927(20000601)21:9<542::AID-MARC542>3.0.CO;2-7.

Fernández-Arrojo, L., Guazzaroni, M.-E., López-Cortés, N., Beloqui, A., and Ferrer, M. (2010). Metagenomic era for biocatalyst identification. *Current Opinion in Biotechnology* 21(6)**,** 725-733. doi: https://doi.org/10.1016/j.copbio.2010.09.006.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A.P.M.D., et al. (2005). Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology* 7(12)**,** 1996-2010. doi: 10.1111/j.1462-2920.2005.00920.x.

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W.R., Golyshina, O.V., and Golyshin, P.N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial Biotechnology* 9(1)**,** 22-34. doi: doi:10.1111/1751-7915.12309.

Foster, Z.S.L., Sharpton, T.J., and Grunwald, N.J. (2017). Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *Plos Computational Biology* 13(2). doi: 10.1371/journal.pcbi.1005404.

Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X.C., et al. (2015). Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology* 13**,** 360. doi: 10.1038/nrmicro3451.

Gharechahi, J., and Salekdeh, G.H. (2018). A metagenomic analysis of the camel rumen's microbiome identifies the major microbes responsible for lignocellulose degradation and fermentation. *Biotechnology for Biofuels* 11(1)**,** 216. doi: 10.1186/s13068-018-1214-9.

Gomez-Alvarez, V., Teal, T.K., and Schmidt, T.M. (2009). Systematic artifacts in metagenomes from complex microbial communities. *The Isme Journal* 3**,** 1314. doi: 10.1038/ismej.2009.72.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., et al. (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* 331(6016)**,** 463-467. doi: 10.1126/science.1200387.

Holtshausen, L., Schwartzkopf-Genswein, K.S., and Beauchemin, K.A. (2013). Short Communication: Ruminal pH profile and feeding behaviour of feedlot cattle transitioning from a high-forage to a high-concentrate diet. *Canadian Journal of Animal Science* 93(4)**,** 529-533. doi: 10.4141/cjas2013-073.

Jones, D.R., Uddin, M.S., Gruninger, R.J., Pham, T.T.M., Thomas, D., Boraston, A.B., et al. (2017). Discovery and characterization of family 39 glycoside hydrolases from rumen anaerobic fungi with polyspecific activity on rare arabinosyl substrates. *Journal of Biological Chemistry* 292(30)**,** 12606-12620. doi: 10.1074/jbc.M117.789008.

Joshi, M.D., Sidhu, G., Pot, I., Brayer, G.D., Withers, S.G., and McIntosh, L.P. (2000). Hydrogen bonding and catalysis: a novel explanation for how a single amino acid

substitution can change the ph optimum of a glycosidase11Edited by M. F. Summers. *Journal of Molecular Biology* 299(1), 255-279. doi: https://doi.org/10.1006/jmbi.2000.3722.
Kaminski, J., Gibson, M.K., Franzosa, E.A., Segata, N., Dantas, G., and Huttenhower, C. (2015). High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED. *Plos Computational Biology* 11(12). doi: 10.1371/journal.pcbi.1004557.
Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics (Oxford, England)* 28(24), 3211-3217. doi: 10.1093/bioinformatics/bts611.
Levin, B.J., Huang, Y.Y., Peck, S.C., Wei, Y., Campo, A.M.D., Marks, J.A., et al. (2017). A prominent glycyl radical enzyme in human gut microbiomes metabolizes trans-4-hydroxy-L-proline. *Science* 355(6325), 595-+. doi: 10.1126/science.aai8386.
Li, D.H., Liu, C.M., Luo, R.B., Sadakane, K., and Lam, T.W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31(10), 1674-1676. doi: 10.1093/bioinformatics/btv033.
Li, F., and Guan, L.L. (2017). Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Applied And Environmental Microbiology*. doi: 10.1128/AEM.00061-17.
Li, Y., and Zhang, Y. (2009). REMO: A New Protocol to Refine Full Atomic Protein Models from C-alpha Traces by Optimizing Hydrogen-Bonding Networks. *Proteins* 76(3), 665-676. doi: 10.1002/prot.22380.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42(D1), D490-D495. doi: 10.1093/nar/gkt1178.
McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* 40(10), 4288-4297. doi: 10.1093/nar/gks042.
McNeil, M., Darvill, A.G., Fry, S.C., and Albersheim, P. (1984). Structure and Function of the Primary Cell Walls of Plants. *Annual Review of Biochemistry* 53(1), 625-663. doi: 10.1146/annurev.bi.53.070184.003205.
Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., et al. (2008a). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9(1), 386. doi: 10.1186/1471-2105-9-386.
Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., et al. (2008b). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *Bmc Bioinformatics* 9. doi: 10.1186/1471-2105-9-386.
Montanholi, Y.R., Swanson, K.C., Palme, R., Schenkel, F.S., McBride, B.W., Lu, D., et al. (2010). Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. *Animal* 4(5), 692-701. doi: 10.1017/s1751731109991522.
Munoz-Munoz, J., Cartmell, A., Terrapon, N., Henrissat, B., and Gilbert, H.J. (2017). Unusual active site location and catalytic apparatus in a glycoside hydrolase

family. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1701130114.

Neves, A.L.A., Li, F., Ghoshal, B., McAllister, T., and Guan, L.L. (2017). Enhancing the Resolution of Rumen Microbial Classification from Metatranscriptomic Data Using Kraken and Mothur. *Frontiers in Microbiology* 8(2445). doi: 10.3389/fmicb.2017.02445.

Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., et al. (2011). Snapshot of the Eukaryotic Gene Expression in Muskoxen Rumen-A Metatranscriptomic Approach. *Plos One* 6(5). doi: 10.1371/journal.pone.0020521.

Ricard, G., McEwan, N.R., Dutilh, B.E., Jouany, J.-P., Macheboeuf, D., Mitsumori, M., et al. (2006). Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. *BMC Genomics* 7**,** 22-22. doi: 10.1186/1471-2164-7-22.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1)**,** 139-140. doi: 10.1093/bioinformatics/btp616.

Robinson, M.D., and Smyth, G.K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* 9(2)**,** 321-332. doi: 10.1093/biostatistics/kxm030.

Russell, J.B., and Rychlik, J.L. (2001). Factors that alter rumen microbial ecology. *Science* 292(5519)**,** 1119-1122. doi: 10.1126/science.1058830.

Schnoes, A.M., Brown, S.D., Dodevski, I., and Babbitt, P.C. (2009). Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLOS Computational Biology* 5(12)**,** e1000605. doi: 10.1371/journal.pcbi.1000605.

Seshadri, R., Leahy, S.C., Attwood, G.T., Teh, K.H., Lambie, S.C., Cookson, A.L., et al. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nature Biotechnology* 36(4)**,** 359-367. doi: 10.1038/nbt.4110.

Shaani, Y., Zehavi, T., Eyal, S., Miron, J., and Mizrahi, I. (2018). Microbiome niche modification drives diurnal rumen community assembly, overpowering individual variability and diet effects. *The ISME Journal*. doi: 10.1038/s41396-018-0203-0.

Sowa, Y., and Berry, R.M. (2008). Bacterial flagellar motor. *Quarterly Reviews of Biophysics* 41(2)**,** 103-132. doi: 10.1017/S0033583508004691.

Stephen, A.M. (1983). *The Polysaccharides.* San Diego, CA, USA: Academic.

Stuart, D.I., Courcelle, E., M √ ©toz, F., and Gouet, P. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15(4)**,** 305-308. doi: 10.1093/bioinformatics/15.4.305.

Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J., et al. (2011). "The Complete Genome Sequence of Fibrobacter succinogenes S85 Reveals a Cellulolytic and Metabolic Specialist". (United States, North America: Public Library of Science).

Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B., Wu, C.H., and the UniProt, C. (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31(6)**,** 926-932. doi: 10.1093/bioinformatics/btu739.

Svartström, O., Alneberg, J., Terrapon, N., Lombard, V., de Bruijn, I., Malmsten, J., et al. (2017). Ninety-nine de novo assembled genomes from the moose (Alces alces) rumen microbiome provide new insights into microbial plant biomass degradation. *The ISME Journal*. doi: 10.1038/ismej.2017.108.

Thompson, S. (2015). *The effect of diet type on residual feed intake and the use of infrared thermography as a method to predict efficiency in beef bulls.* M.Sc., University of Manitoba.

Törrönen, A., Harkki, A., and Rouvinen, J. (1994). Three-dimensional structure of endo-1,4-beta-xylanase II from Trichoderma reesei: two conformational states in the active site. *The EMBO Journal* 13(11)**,** 2493-2501.

Vardakou, M., Dumon, C., Murray, J.W., Christakopoulos, P., Weiner, D.P., Juge, N., et al. (2008). Understanding the Structural Basis for Substrate and Inhibitor Recognition in Eukaryotic GH11 Xylanases. *Journal of Molecular Biology* 375(5)**,** 1293-1305. doi: https://doi.org/10.1016/j.jmb.2007.11.007.

Wang, L., Hatem, A., Catalyurek, U.V., Morrison, M., and Yu, Z. (2013). Metagenomic Insights into the Carbohydrate-Active Enzymes Carried by the Microorganisms Adhering to Solid Digesta in the Rumen of Cows. *PLoS ONE* 8(11)**,** 1-11. doi: 10.1371/journal.pone.0078507.

Weimer, P.J. (1998). Manipulating ruminal fermentation: a microbial ecological perspective. *Journal of Animal Science* 76(12)**,** 3114-3122. doi: 10.2527/1998.76123114x.

Weimer, P.J. (2015). Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. *Frontiers in Microbiology* 6(296). doi: 10.3389/fmicb.2015.00296.

Wilkens, C., Busk, P.K., Pilgaard, B., Zhang, W.-J., Nielsen, K.L., Nielsen, P.H., et al. (2017). Diversity of microbial carbohydrate-active enzymes in Danish anaerobic digesters fed with wastewater treatment sludge. *Biotechnology for Biofuels* 10(1)**,** 158. doi: 10.1186/s13068-017-0840-y.

Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15(3)**,** R46-R46. doi: 10.1186/gb-2014-15-3-r46.

Wu, S., and Zhang, Y. (2007). LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research* 35(10)**,** 3375-3382. doi: 10.1093/nar/gkm251.

Xu, B., Xu, W., Li, J., Dai, L., Xiong, C., Tang, X., et al. (2015). Metagenomic analysis of the Rhinopithecus bieti fecal microbiome reveals a broad diversity of bacterial and glycoside hydrolase profiles related to lignocellulose degradation. *BMC Genomics* 16(1)**,** 174. doi: 10.1186/s12864-015-1378-7.

Yang, J., Roy, A., and Zhang, Y. (2013). Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* 29(20)**,** 2588-2595. doi: 10.1093/bioinformatics/btt447.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2014). The I-TASSER Suite: protein structure and function prediction. *Nature Methods* 12**,** 7. doi: 10.1038/nmeth.3213.

Yin, Y.B., Mao, X.Z., Yang, J.C., Chen, X., Mao, F.L., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 40(W1)**,** W445-W451. doi: 10.1093/nar/gks479.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9(1)**,** 40. doi: 10.1186/1471-2105-9-40.

5.7 Tables and figures

Table 5.1. Kinetic parameters characterizing the thermal inactivation of Xylanase 1

Temperature (°C)	K_d (min^{-1})	$t_{1/2}$ (min)	D (min)	ΔH^o ($kJ \cdot mol^{-1}$)	ΔG^o ($kJ \cdot mol^{-1}$)	ΔS^o ($J \cdot mol^{-1} \cdot K^{-1}$)
25	0.0017	407.73	1354.46	150.52	78.64	0.24
40	0.0180	38.51	127.92	150.40	76.58	0.24
50	0.2400	2.89	9.59	150.31	72.16	0.24
60	0.92	0.75	2.50	150.23	70.75	0.24

K_d, inactivation rate constant; $t_{1/2}$, half-time (i.e., the time after which activity is reduced to one-half of the initial value); D, the time required to reduce enzymatic activity to 10% of its orginal value; ΔH^o, activation enthalpy; ΔG^o, activation free-energy barrier; ΔS^o, activation entropy of thermal denaturation.

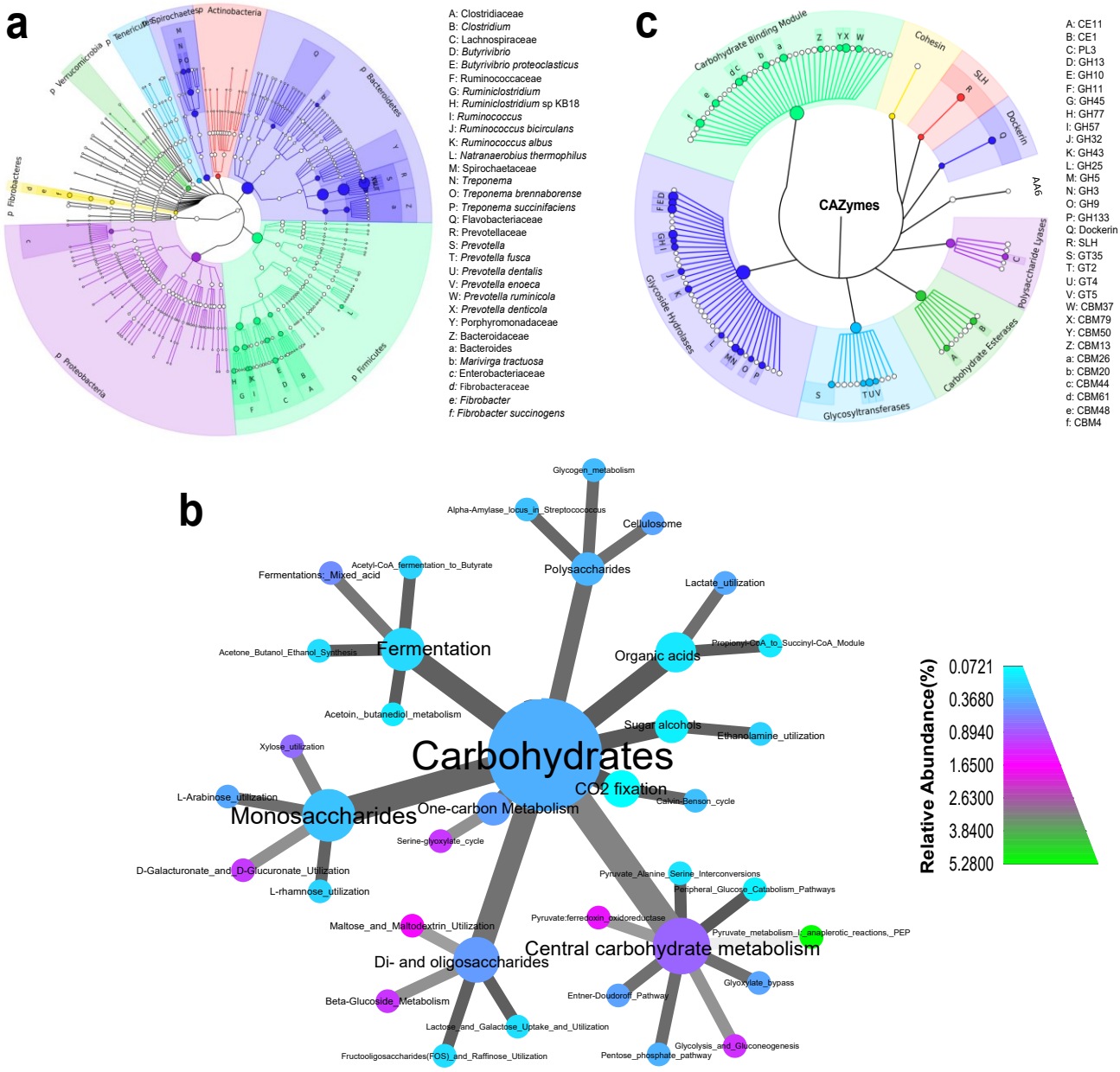


Figure 5.1. Taxonomic composition and microbial gene families obtained from metatranscriptome data in beef cattle. **a** Taxonomic cladogram showing the most abundant detected taxa (relative abundance ≥ 0.1% in at least half of the samples). The six rings of the cladogram stand for phylum (innermost), class, order, family, genus, and species (outermost), respectively. The sizes of the circles indicate the mean average abundance of each taxon. **b** Heat tree displaying the functional capability of the rumen

microbiota. Each node represents the functional categories (up to three levels) and the edges determine where each node fits in the functional hierarchy. Node diameter (and colors) indicate the relative abundance of the functions at level 3. **c** Cladogram showing the most abundant CAZymes. The sizes of the circles indicate the mean average abundance of each CAZyme family.

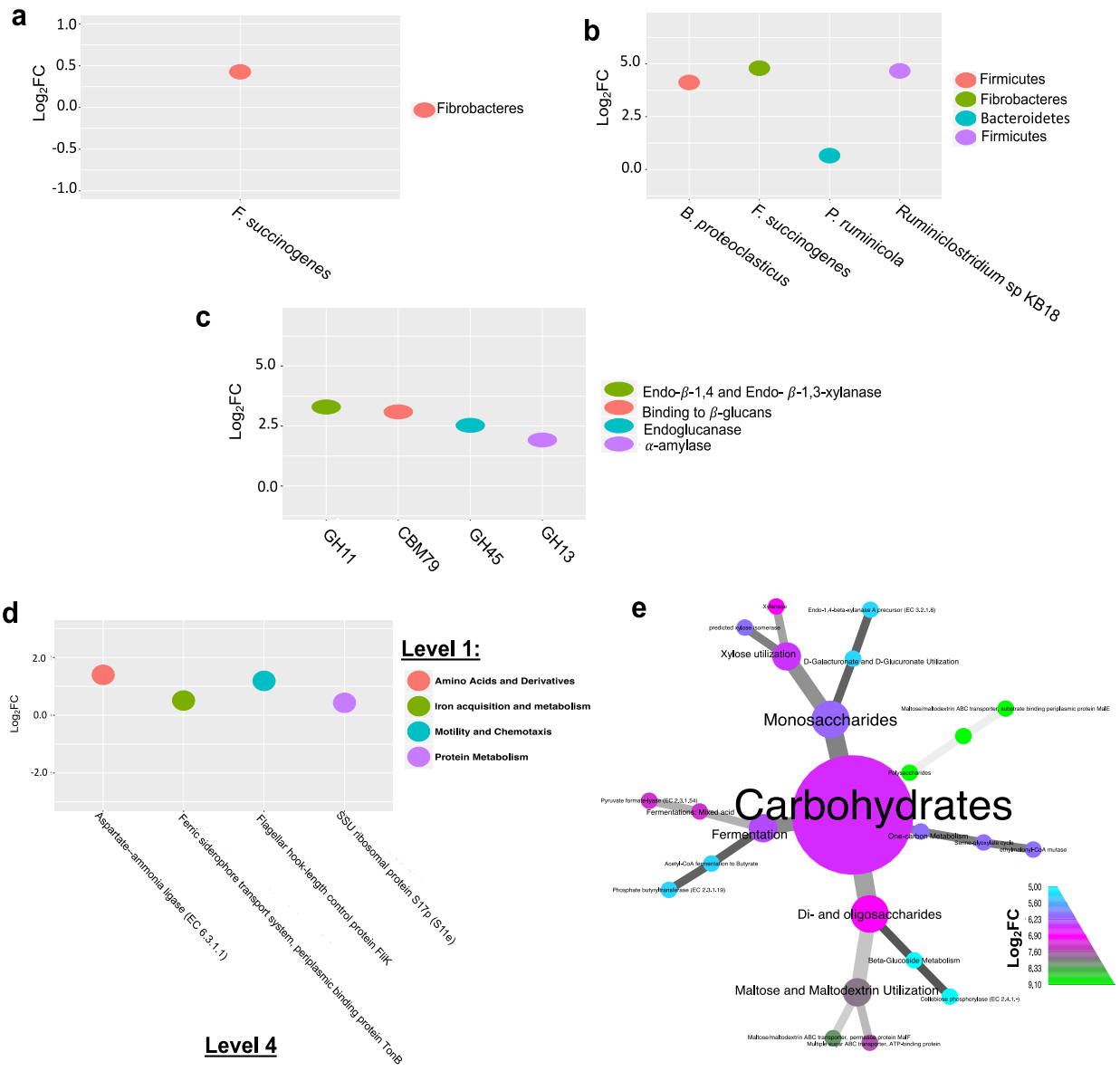


Figure 5.2. Differentially abundant bacterial species (**a-b**), CAZymes (**c**), and microbial functions (**d-e**) detected in the rumen microbiome of feed efficient cattle. Features were significant (P < 0.05) between L-FCR and H-FCR by the trimmed mean of M-values method implemented in *edgeR*. In the heat tree, node diameter (and colors) indicate the log_2fc in the functional categories at level 4.

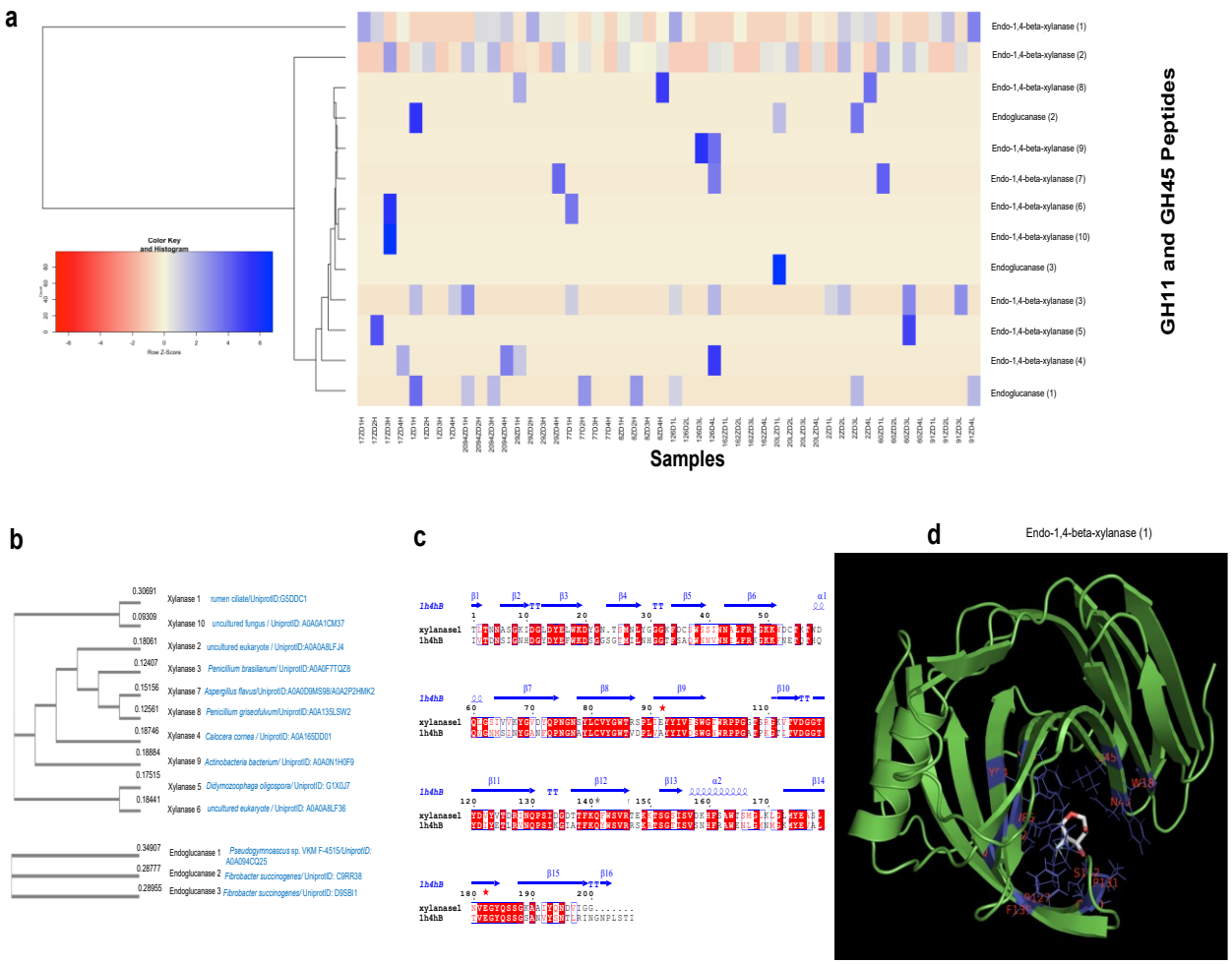


Figure 5.3. Abundance of CAZyme members in the rumen microbiome. **a** Heatmap showing the abundance and distribution of the 13 most abundant members of the GH11 and GH45 families quantified according to ShortBRED (Kaminski et al. 2015). **b** Phylogenetic tree of xylanases and endoglucanases detected in the rumen metatranscriptome generated by the neighbor-joining method. **c** Sequence alignment of xylanase 1 with *1h4hB,* with the residues involved in substrate-binding labeled with a red star as shown by ESPript (Stuart et al. 1999). **d** Theoretical 3D structure of xylanase 1 bound to 1,2-Deoxy-2-Fluoro-Xylopyranose (DFX) (PDB ID: 1c5iA) generated by I-TASSER (Zhang 2008, Yang et al. 2014).

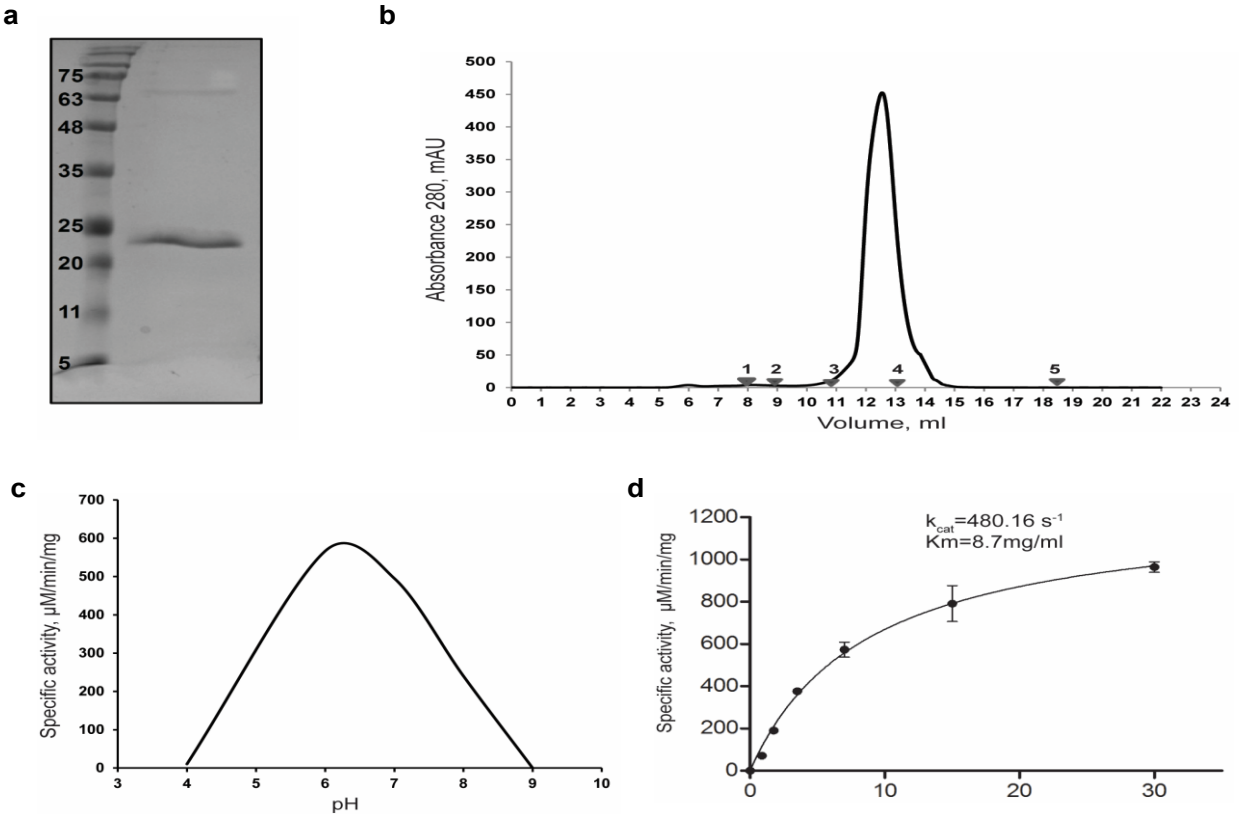


Figure 5.4. Purification and characterization of oligomeric state, and kinetic analysis of Xylanase 1. SDS-PAGE analysis (**a**), oligomeric state characterization (**b**), pH curve (**c**) and Michaelis-Menten plot (**d**) of Xylanase 1**.**

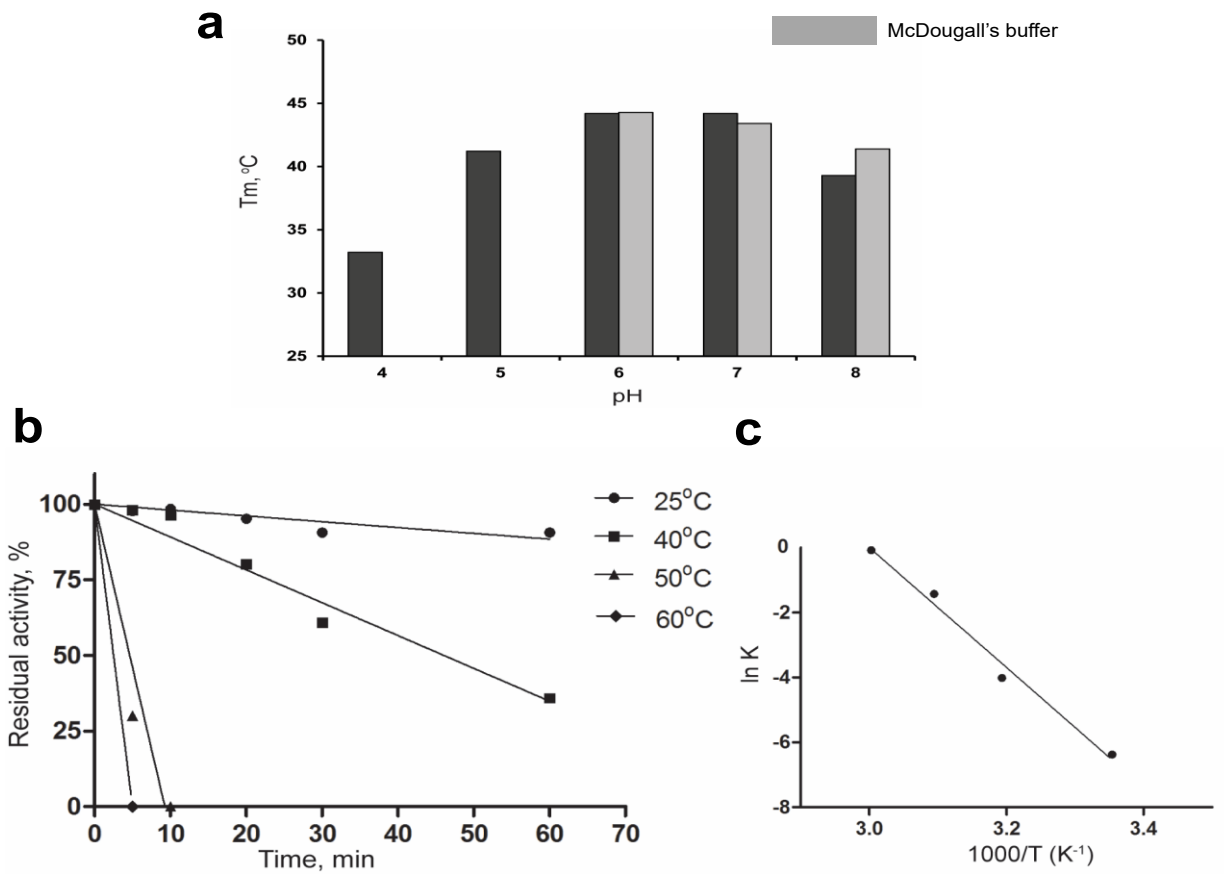


Figure 5.5. Thermal stability of Xylanase 1. T_m determined by differential scanning fluorimetry in different buffers (**a**) and residual activity (%) of Xylanase 1 at different temperatures (**b**) with the respective Arrhenius plot (**c**).

Chapter 6

General Discussion

6.1 Significance of the research

The advance in omics technologies has led to tremendous progress in our understanding of the rumen microbiome and its influence on host feed efficiency. However, significant gaps remain in the literature concerning the interactions between host variation in feed efficiency and the dynamics of rumen microbial populations. Channeling efforts towards investigating these interactions offers an opportunity to enhance cattle productivity, since the rumen microbiota plays a significant role in the host feed efficiency. While steps forward have been made in this regard (Belanche et al. 2012, Fernando et al. 2010, Roehe et al. 2016), little is known about the driving forces that influence the relationship between the rumen microbiota and host individual variation, and how their interactive effects on animal productivity contribute to the identification of cattle with improved feed efficiency. To address this, Chapter 2 of this thesis examined the interactions between host variation in feed efficiency and the dynamics of the rumen microbial population, contributing to a broader effort to enhance the rumen function and improve cattle feed efficiency.

An intrinsic relationship exists between feed efficiency and microbial functions in the rumen, as revealed by total RNA-based techniques recently developed to investigate the rumen metatranscriptome (Fuyong Li and Le Luo Guan 2017). However, this methodology may be limited as total rRNA-based metatranscriptomics may underestimate the abundance of lowly expressed transcripts (Li et al. 2019a). Thus, in-depth analysis of the associations between rumen taxonomic profiles and feed efficiency using other approaches (such as mRNA-based methods) are urgently needed in order to enhance the resolution of the rumen microbial classification. Chapter 3 filled this knowledge gap and assessed the rumen microbial classification from metatranscriptomic data using mRNA-based methods, improving our understanding of the linkage between rumen taxonomic profiles and feed efficiency.

Although the role of the rumen microbiome in the host feed efficiency has been widely investigated using metatranscriptomic analysis (Fuyong Li and Le Luo Guan 2017, Comtet-Marre et al. 2017), little is known regarding variations in microbial signatures that could be targeted to improve feed efficiency. Chapter 4 employed statistical approaches that allowed the identification of breed-specific microbial signatures by capturing within- and between-individual variation observed in closely related breeds. The findings presented in Chapter 4 suggested that the identification of rumen microbial signatures in genetically similar breeds may benefit cattle production by providing further possibilities to target specific taxa and functions and to discover enzymes that can be used to maximize feed digestion in the rumen.

The rumen is a potent reservoir of microbial enzymes and the role of these biocatalysts (CAZymes) in the feed digestion has been extensively studied using omics technologies in order to discover new enzymes with applications in the biotechnology industry (Hess et al. 2011, Meng Qi et al. 2011, Seshadri et al. 2018). However, the data described in the literature are limited to describe CAZyme family classifications and their relative abundance in the rumen, and albeit this information is necessary, it does not give details on the functions of distinct members of these families and their association with host feed efficiency. To address this gap, Chapter 5 used a targeted functional profiling approach to identify distinct members of CAZyme families of interest and pinpointed enzymes of unknown functions from their abundance, enzymatic activity, and distribution in the rumen of feed efficient cattle. Overall, the data presented in this thesis contribute to our fundamental understanding concerning the role of the rumen microbiome in cattle feed efficiency (Chapters 2, 3 and 4) and provide opportunities to further explore the potential of the rumen as a reservoir for novel enzyme discovery (Chapter 5).

6.2 Understanding the interactions between host variation in feed efficiency and the dynamics of the rumen microbial population

Understanding the interactions between the host and the microbiota is crucial for predicting microbial shifts and identifying individuals that are either responsive or resilient to dietary changes (Bashiardes et al. 2018). Most rumen studies to date have generated results that are usually qualitative (Belanche et al. 2011) or semi-quantitative (Comtet-

Marre et al. 2017), but not quantitative, which may not be sufficient to capture the dynamic changes of the rumen microbiota according to the host responsiveness to the diet. Defining the “baseline” of the quantified rumen microbiota in cattle undergoing dietary changes (Chapter 2) revealed that animals grouped according to the magnitude of rumen microbial changes reflected the individual responses to the dietary interventions. To our knowledge, this is the first study that validated the relationship between feed efficiency and host individual variation in the rumen microbial abundance following the dietary switches.

The interest in the stratification of animals via the magnitude of change in baseline microbiota following dietary changes is driven by previous observations of variations in performance across individuals, even when maintained on the same diet (Z. P. Li et al. 2016, Brulc et al. 2009, Zhou et al. 2018). The approach used in Chapter 2 to group animals based on the individual variation in the abundance of microbial populations showed a relationship with feed efficiency (measured as FCR). The outcomes of Chapter 2 suggested that individual hosts exhibiting variability in bacterial abundance above or below a certain cut-off ($log_2fc < -1$ or > 1) were more efficient in terms of FCR than those presenting a stable variability in the bacterial abundance. In practice, this approach is likely to be more effective when the diets shift from forage to grain than from grain to forage, as the former sequence of dietary treatments caused a larger variation in microbial baselines between the diets.

Yet, the results presented in Chapter 2 are further limited by the absence of microbial diversity data characterizing each group of log_2fc in rumen microbes, and this should be prioritized in future studies in order to get insight into changes in specific genera/species that may be critical for feed digestion. Moreover, samples were only available at two-time points prior to the determination of the baseline rumen microbiota, and thus future investigations should include more sampling points to clearly define the precise baseline rumen microbiota that is necessary to obtain the stratification of animals undergoing dietary changes. One might also argue that variations in the content of dietary components (e.g., NDF) could influence the effectiveness of microbial fermentation of the feed, and consequently the results found in Chapter 2. In this case, it might be beneficial

to focus future research efforts towards analyzing the effects of different types of dietary fibers on the inter-individual variability of the rumen microbiota in terms of feed degradation, as the type of fibers affects the time of rumination and the microbial fermentation of the feed (NRC 2016). Whatever the limitations might be, these data showed that the dynamics of the rumen microbial population is intimately associated with inter-individual variability in the baseline microbiota, and these findings could be used to design better feeding strategies to enhance the rumen function and to identify cattle with improved feed efficiency.

6.3 Understanding the rumen microbiota using mRNA-based metatranscriptomics

A recent study by Fuyong Li et al. (2016) developed a Mothur (Schloss et al. 2009) based pipeline to assess active rumen microbiota in total RNA sequencing datasets. Despite the efficacy of this pipeline to investigate linkages between the active rumen microbiome (structure and function) and feed efficiency in beef cattle (Fuyong Li and Le Luo Guan 2017), it still remained a challenge for researchers to determine which approach (total RNA or mRNA-based methods) of taxonomic classification delivered the most realistic representation of the rumen microbial community. In chapter 3, we adapted the Kraken pipeline (Wood and Salzberg 2014) to analyze mRNA sequences via reference genomes of rumen microorganisms. The comparative analysis of Mothur and Kraken revealed that both pipelines showed the rumen being dominated by *Prevotella*, *Treponema*, *Ruminoccocus*, *Fibrobacter*, and *Butyrivibrio*, which are considered as part of a “core bacterial microbiome” (Henderson et al. 2015). However, the Kraken pipeline classified rumen microbes beyond the genus level when compared with Mothur, and this resulted in additional information of the species and their functions in the rumen.

Overall, the Kraken based approach generated a higher resolution of the rumen microbiota because the reference database used to annotate each microbial sequence to the lowest common ancestral (Wood and Salzberg 2014) was built based on all known microbial genomes present in the NCBI database (Neves et al. 2017). Although Kraken enhanced the microbial classification at the species level, identification of archaea in the rumen was challenging due to a lack of archaeal reference genomes for the rumen microbiome. These issues highlight the importance of further strengthening the Kraken

database through the inclusion of more genomes to enable a more accurate classification. Another issue the needs to be addressed is the high number of sequences unclassified (65%) by Kraken, which points to a need for further research on poorly studied microbial groups. Two possible causes for the high percentage of unclassified sequences were a) the limited number of rumen species deposited in JGI at the time of the data analysis and b) the lack of sequencing depth of the total RNA approach used in the study. Thus, there is a need to include new genomes and/or metagenome assembled genomes from the rumen microbiome in the customized Kraken database as well as a deeper sequence depth if we use mRNA to improve the results of the data analysis in future studies.. Regardless of the approach undertaken in the future, the only way for improvement is through continued strengthening of the databases by including additional information of whole-genome sequencing of rumen isolates, as the ability to culture rumen microorganisms is still limited.

6.4 Associations between microbial signatures, feed efficiency and host genetics

Chapter 4 built on the bioinformatic pipeline developed as part of Chapter 3. Taking advantage of the analytical capabilities outlined in that previous work, we determined the taxonomic profiles of the rumen microbiome of Angus cattle, which are known for their superior performance and beef quality (McLean and Schmutz 2009, Wolfger et al. 2016). The animal effects that can impact feed efficiency include feeding behavior, energy metabolism, the rumen microbiota, and the genetic background of the host (Cantalapiedra-Hijar et al. 2018). Although these factors are usually interconnected with each other, research questions regarding the associations between the genetic background of the host with the rumen microbiota as well as with feed efficiency have raised much interest in recent years (Roehe et al. 2016). One of these questions is to address the current lack of information regarding the contributions of the genetic makeup of the host (e.g., genetically similar breeds of Black and Red Angus) to variations in specific microbial signatures that could be targeted to improve ruminal function and feed efficiency. Black and Red Angus cattle were the chosen candidates to study variations in microbial signatures in Chapter 4 because they are genetically similar and comparisons between them with respect to the taxonomic and functional profiles could help uncover

microbial signatures related to feed efficiency of the Angus breed, which comprises a large proportion of the beef cattle population in North America.

In Chapter 4, the functional potential of the rumen microbiota was investigated through innovative statistical models (sPLS-DA and ANCOM) to account for the compositional aspect of the microbiome data and to discover microbial signatures that characterized the closely related breeds of Black and Red Angus. To the best of our knowledge, these are the first data to provide a further understanding of the relationships that exist between breed-specific microbial signatures (bacterial species and microbial functions) and host phenotype (Black vs. Red Angus) at the RNA-level. Despite the close genetic similarities between Red and Black Angus, Chapter 4 revealed unique bacterial phylotypes and functions that differentiated the rumen microbiome of each breed. It was found that rumen fibrolytic species and a diverse set of microbial gene families were more abundant in Black Angus, whereas poorly-characterized bacterial species and specific microbial genes (e.g., carbohydrate pathways) were more predominant in the rumen of Red Angus.

With its limited sample size, however, this study lacked the power necessary to find differences in feed efficiency (FCR) between Black and Red Angus. Thus, care should be taken while interpreting the data outlined in Chapter 4 as the power of this study may not support the relatively strong statement of 'genetic effects/influences on the rumen microbiome'. Power analysis showed that the minimal sample size to detect a FCR effect size of 1.5 in samples collected from Black and Red Angus over four-time points, with a power of 0.9 and p-value <0.5 is of 9 animals per breed (Figure S5), indicating that this number of replicates needs to be included in future experiments to get more accurate results. Despite these issues, it was revealed associations between microbial signatures (*Chitinophaga pinensis* and *Clostridium stercorarium*) and FCR adjusted to the sampling time across breeds, suggesting that these bacterial species may play a key role in the feed conversion efficiency of forage-fed bulls. Further studies with the sample size mentioned previously might offer a deeper insight into the microbial signatures that are related to host genetics and feed efficiency. Moreover, the inclusion of other breeds might be beneficial to elucidate the host genetic influence on the rumen microbiome. Finally,

the microbial signatures identified in Chapter 3 should be validated using other technologies (such as qPCR) to confirm their contribution to growth performance in Angus cattle reared in other geographical locations and exhibiting different feed efficiencies. In this context, it was attempted to design primers to target the microbial signatures identified in this study (e.g., *Chitinophaga pinensis* and *Clostridium stercorarium*). However, the primers were not successful in amplifying the targeted species due to the lack of representative isolates of the rumen microbiota that could be used as templates to test the specificity of the primer pairs for those species.

6.5 Discovery of novel CAZymes in the rumen through selective pressure on the native microbial population and targeted functional profiling

The idea of discovering CAZymes through selective pressure on the rumen native microbial population and targeted functional profiling (Chapter 5) was conceptualized from the relationship that exists between the rumen microbiota and host feed efficiency (Chapters 2 and 4). The starting point for the development of this strategy was based on the hypothesis that lignocellulolytic microbes (and their encoded enzymes) are more abundant in the rumen of feed efficient cattle compared to inefficient animals fed the same diet. Previous research supported that hypothesis by showing that selective pressure to enrich the rumen microbial environment for the desired biocatalyst could be achieved when specific substrates are included in the diet in order to favor the growth of microbes specialized at expressing the enzymes of interest (Weimer 1998). This concept was incorporated in Chapter 5, and the data showed that the relative abundance of bacterial species was altered after the long-term forage feeding used to enrich the rumen microbiome for lignocellulolytic microbes according to the groups of feed efficiency (L-FCR and H-FCR). Of the 115 species analyzed, *Fibrobacter succinogens,* a cellulose degrader species (Suen et al. 2011), was the only species affected by the feed efficiency groups; it showed a nearly 0.5-$\log_2$-fold change increase ($P < 0.05$) in L-FCR relative to H-FCR. Our results also showed that the relative abundance of *Fibrobacter succinogens* and other common plant cell wall degraders (*Butyrivibrio proteoclasticus* and *Ruminiclostridium* sp KB18) exhibited > 4-$\log_2$-fold change increase on the 180th day relative to day 0 in L-FCR compared to H-FCR, confirming the enrichment of the rumen

microbiome for lignocellulolytic microbes after the long-term forage feeding. To further investigate whether feed efficiency affects CAZymes, we then analyzed all CAZymes families in our dataset in relation to the FCR groups. Chapter 5 showed that GH11 (endo-β-1,4-xylanase - EC 3.2.1.8), GH45 (endoglucanase: EC 3.2.1.4), and CBMs connected to cellulose degradation (CBM79) exhibited >2.5-$\log_2$-fold change increase in their relative abundance in L-FCR compared to H-FCR after the long-term feeding. Taken together, our results indicated that the use of a single type of fibrous diets promotes the increase in the relative abundance of fibrolytic microbes and lignocellulolytic enzymes in the rumen of feed efficient cattle over time.

An exciting aspect of the findings outlined in Chapter 5 is that they are not limited in classifying CAZyme families and determining their abundance in the rumen as observed in previous research (Fuyong Li and Le Luo Guan 2017, Hess et al. 2011, Meng Qi et al. 2011). Instead, it showed the importance of assessing individual members of CAZyme families to guide the discovery of new enzymes and get a comprehensive picture of their ecological distribution in the rumen microbiome. This is particularly relevant by the fact that individual proteins within a protein family can have vastly diverse functions (Franzosa et al. 2015, Schnoes et al. 2009, Levin et al. 2017), and this information may be ignored when the focus is only on the identification of the protein families in the microbiome. I recognize that the identification of CAZyme families followed by the screening of genes encoding novel CAZymes has led to the discovery of thousands of putative enzymes in metagenomic or metatranscriptomic datasets (Brulc et al. 2009, Hess et al. 2011, Wang et al. 2013, Svartström et al. 2017, Gharechahi and Salekdeh 2018, Meng Qi et al. 2011, Comtet-Marre et al. 2017, Fuyong Li and Le Luo Guan 2017), but then a question arises: “why only a handful of these candidate enzymes have entered industrial application?”. This is a serious issue that needs to be carefully considered by the researchers interested in exploring the potential of microbiomes for enzyme discovery. I came to the conclusion that the problem stems from annotation mistakes (sequence/activity incoherence) in publicly databases (Fernández-Arrojo et al. 2010, Ferrer et al. 2016). Actually, the literature cites many of these examples like the one reported by Jiménez et al. (2012) who discovered a novel cold-tolerant esterase using metagenomics approaches, but this protein was annotated in the database as a *MarR*

family transcriptional regulator! These outcomes indicate that database entries are not sufficiently reliable.

In Chapter 5, the targeted functional profiling method (Kaminski et al. 2015) was applied to screen members of CAZyme families against a *de novo* protein reference database and profile their abundance and distribution in the rumen microbiome without relying on the annotations of external databases. By screening 775 uncharacterized members of the family GH11 and 409 of the family GH45, Chapter 5 identified 18 putative xylanases (GH11) and three putative endoglucanases (GH45). Although these enzymes act on lignocellulosic substrates (xylose and cellulose), they exhibited different abundances and distributions in the rumen metatranscriptome, suggesting that they may perform distinct activities within the GH11 and GH45 families. Later, it was investigated key active sites and the tertiary conformation of GH11 and GH45 enzymes and prioritized the enzymatic characterization of xylanase 1 (the most broadly distributed and abundant xylanase in our rumen metatranscriptome dataset). Several experiments involving cloning and protein expression and purification of xylanase 1, as well as differential scanning fluorimetry, size-exclusion chromatography, kinetic measurements, and thermodynamic calculations were performed to characterize that enzyme. The biochemical proof of the xylanolytic activity of the most locally abundant and widespread member of GH11 family using homology modeling and enzymatic assays have demonstrated the usefulness of the strategy outlined in Chapter 5 and strengthened our knowledge of the role played by this previously uncharacterized enzyme. However, further studies are needed to elucidate the crystal structure of xylanase 1 and test the enzymatic activity of all putative xylanases and endoglucanases identified in Chapter 5.

6.6 Implications and future directions

Despite the limitations described above, this thesis provides fundamental knowledge concerning the role of the rumen microbiome in cattle feed efficiency and offers opportunities to further explore the potential of the rumen as a source for novel enzyme discovery. Chapter 2 demonstrated how the interactions between the diet and host variation in the abundance of rumen microbes could be used to assess feed efficiency in ruminants, and these data could be beneficial to animal husbandry in two ways. First, by

understanding the dynamics of the ruminal microbial population in response to dietary changes, researchers will be able to design better feeding strategies to improve the rumen function and host performance, since the microbial community plays an important role in the digestion of feedstuffs (Belanche et al. 2012, Fernando et al. 2010, Roehe et al. 2016). Second, our approach could serve as a cheap strategy to quickly assess rumen microbial shifts in cattle populations undergoing dietary changes under field situations (e. g., feedlots), and to identify feed efficient or inefficient animals under the same dietary management since we found a relationship between FCR and the abundance of rumen microorganisms. The animals could then be managed differently for the most profitable way to the farmers. However, the data presented in Chapter 2 are limited by the lack of microbial diversity information for each group of log2fc in rumen microbes, and this should be prioritized in future experiments as discussed previously.

The studies outlined in Chapters 3 and 4 developed a pipeline to explore the taxonomic and functional characteristics of the rumen microbiome and applied novel statistical approaches as alternatives to the traditional analytical methods discussed in Chapter 1. The pipeline described in Chapter 3 will offer opportunities to analyze metatranscriptomic sequence data (~140 bp in length) of rumen samples by providing further possibilities to enhance the resolution of the rumen microbial classification. Chapter 4, in turn, showed results on the link between the rumen microbiome, breed differences, and FCR and is of interest for the scientific community because of the statistical methodologies that were applied in combination with metatranscriptomics. The future directions for Chapter 3 include the inclusion of 913 bacterial and archaeal genomes (Stewart et al. 2018) in the Kraken database and the development of a pipeline combining the approaches of Fuyong Li et al. (2016) and Neves et al. (2017) for a more inclusive and representative classification of the rumen microbiome. For Chapter 4, the main point to be considered in future experiments includes improvements in the design of the study to enhance the power of the analysis and the reproducibility of the results.

Chapter 5 developed a strategy to discover new enzymes through the application of selective pressure on the rumen native population and the targeted functional profiling of CAZyme families. This strategy can uncover previously unappreciated enzymes and

this might offer further possibilities for meeting increasing market demands for novel biocatalysts with applications in animal feed industries. An interesting aspect to be explored in the future using the strategy outlined in Chapter 5 is the implementation of a feeding intervention that includes recalcitrant diets (e.g., sugar cane bagasse) at a rate as high as possible without compromising the animal health. This approach will encourage the evolution of microbes specialized at digesting lignocellulose, and it is likely that many other lignocellulolytic enzymes with high efficiency will be discovered through this strategy. In addition, such an approach may also be applied to discover new enzymes with biotechnology implications such as biofuels.

6.7 Literature cited

Bashiardes, S., Godneva, A., Elinav, E., and Segal, E. (2018). Towards utilization of the human genome and microbiome for personalized nutrition. *Current Opinion in Biotechnology* 51**,** 57-63. doi: 10.1016/j.copbio.2017.11.013.

Belanche, A., Abecia, L., Holtrop, G., Guada, J.A., Castrillo, C., de la Fuente, G., et al. (2011). Study of the effect of presence or absence of protozoa on rumen fermentation and microbial protein contribution to the chyme1. *Journal of Animal Science* 89(12)**,** 4163-4174. doi: 10.2527/jas.2010-3703.

Belanche, A., Doreau, M., Edwards, J.E., Moorby, J.M., Pinloche, E., and Newbold, C.J. (2012). Shifts in the Rumen Microbiota Due to the Type of Carbohydrate and Level of Protein Ingested by Dairy Cattle Are Associated with Changes in Rumen Fermentation. *Journal of Nutrition* 142(9)**,** 1684-1692. doi: 10.3945/jn.112.159574.

Brulc, J.M., Antonopoulos, D.A., Berg Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., et al. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences* 106(6)**,** 1948-1953. doi: 10.1073/pnas.0806191105.

Comtet-Marre, S., Parisot, N., Lepercq, P., Chaucheyras-Durand, F., Mosoni, P., Peyretaillade, E., et al. (2017). Metatranscriptomics Reveals the Active Bacterial and Eukaryotic Fibrolytic Communities in the Rumen of Dairy Cow Fed a Mixed Diet. *Frontiers in Microbiology* 8(67). doi: 10.3389/fmicb.2017.00067.

Fernández-Arrojo, L., Guazzaroni, M.-E., López-Cortés, N., Beloqui, A., and Ferrer, M. (2010). Metagenomic era for biocatalyst identification. *Current Opinion in Biotechnology* 21(6)**,** 725-733. doi: https://doi.org/10.1016/j.copbio.2010.09.006.

Fernando, S.C., Purvis, H.T., II, Najar, F.Z., Sukharnikov, L.O., Krehbiel, C.R., Nagaraja, T.G., et al. (2010). Rumen Microbial Population Dynamics during Adaptation to a High-Grain Diet. *Applied and Environmental Microbiology* 76(22)**,** 7482-7490. doi: 10.1128/aem.00388-10.

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W.R., Golyshina, O.V., and Golyshin, P.N. (2016). Estimating the success of enzyme bioprospecting through

metagenomics: current status and future trends. *Microbial Biotechnology* 9(1), 22-34. doi: doi:10.1111/1751-7915.12309.

Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X.C., et al. (2015). Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology* 13, 360. doi: 10.1038/nrmicro3451.

Gharechahi, J., and Salekdeh, G.H. (2018). A metagenomic analysis of the camel rumen's microbiome identifies the major microbes responsible for lignocellulose degradation and fermentation. *Biotechnology for Biofuels* 11(1), 216. doi: 10.1186/s13068-018-1214-9.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P.H., et al. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports* 6. doi: 10.1038/srep19175.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., et al. (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* 331(6016), 463-467. doi: 10.1126/science.1200387.

Jiménez, D.J., Montaña, J.S., Álvarez, D., and Baena, S. (2012). A novel cold active esterase derived from Colombian high Andean forest soil metagenome. *World Journal of Microbiology and Biotechnology* 28(1), 361-370. doi: 10.1007/s11274-011-0828-x.

Kaminski, J., Gibson, M.K., Franzosa, E.A., Segata, N., Dantas, G., and Huttenhower, C. (2015). High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED. *Plos Computational Biology* 11(12). doi: 10.1371/journal.pcbi.1004557.

Levin, B.J., Huang, Y.Y., Peck, S.C., Wei, Y., Campo, A.M.D., Marks, J.A., et al. (2017). A prominent glycyl radical enzyme in human gut microbiomes metabolizes trans-4-hydroxy-L-proline. *Science* 355(6325), 595-+. doi: 10.1126/science.aai8386.

Li, F., and Guan, L.L. (2017). Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Applied And Environmental Microbiology*. doi: 10.1128/AEM.00061-17.

Li, F., Henderson, G., Sun, X., Cox, F., Janssen, P.H., and Guan, L.L. (2016a). Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches. *Frontiers in Microbiology* 7(987). doi: 10.3389/fmicb.2016.00987.

Li, F., Hitch, T.C.A., Chen, Y., Creevey, C.J., and Guan, L.L. (2019). Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the rumen microbiome and its associations with feed efficiency in beef cattle. *Microbiome* 7(1), 6. doi: 10.1186/s40168-019-0618-5.

Li, Z.P., Wright, A.D.G., Si, H.Z., Wang, X.X., Qian, W.X., Zhang, Z.G., et al. (2016b). Changes in the rumen microbiome and metabolites reveal the effect of host genetics on hybrid crosses. *Environmental Microbiology Reports* 8(6), 1016-1023. doi: 10.1111/1758-2229.12482.

McLean, K.L., and Schmutz, S.M. (2009). Associations of melanocortin 1 receptor genotype with growth and carcass traits in beef cattle. *Canadian Journal of Animal Science* 89(3), 295-300. doi: 10.4141/CJAS08094.

Neves, A.L.A., Li, F., Ghoshal, B., McAllister, T., and Guan, L.L. (2017). Enhancing the Resolution of Rumen Microbial Classification from Metatranscriptomic Data Using Kraken and Mothur. *Frontiers in Microbiology* 8(2445). doi: 10.3389/fmicb.2017.02445.

NRC (2016). *Nutrient Requirements of Beef Cattle.* Washington, D.C., USA: National Academy Press.

Piro, V.C., Matschkowski, M., and Renard, B.Y. (2017). MetaMeta: integrating metagenome analysis tools to improve taxonomic profiling. *Microbiome* 5(1)**,** 101-101. doi: 10.1186/s40168-017-0318-y.

Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., et al. (2011). Snapshot of the Eukaryotic Gene Expression in Muskoxen Rumen-A Metatranscriptomic Approach. *Plos One* 6(5). doi: 10.1371/journal.pone.0020521.

Roehe, R., Dewhurst, R.J., Duthie, C.-A., Rooke, J.A., McKain, N., Ross, D.W., et al. (2016). Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. *Plos Genetics* 12(2). doi: 10.1371/journal.pgen.1005846.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied & Environmental Microbiology* 75(23)**,** 7537-7541. doi: 10.1128/AEM.01541-09.

Schnoes, A.M., Brown, S.D., Dodevski, I., and Babbitt, P.C. (2009). Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLOS Computational Biology* 5(12)**,** e1000605. doi: 10.1371/journal.pcbi.1000605.

Seshadri, R., Leahy, S.C., Attwood, G.T., Teh, K.H., Lambie, S.C., Cookson, A.L., et al. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nature Biotechnology* 36(4)**,** 359-367. doi: 10.1038/nbt.4110.

Stewart, R.D., Warr, A., Watson, M., Auffret, M.D., Dewhurst, R.J., Roehe, R., et al. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nature Communications* 9(1). doi: 10.1038/s41467-018-03317-6.

Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J., et al. (2011). "The Complete Genome Sequence of Fibrobacter succinogenes S85 Reveals a Cellulolytic and Metabolic Specialist". (United States, North America: Public Library of Science).

Svartström, O., Alneberg, J., Terrapon, N., Lombard, V., de Bruijn, I., Malmsten, J., et al. (2017). Ninety-nine de novo assembled genomes from the moose (Alces alces) rumen microbiome provide new insights into microbial plant biomass degradation. *The ISME Journal*. doi: 10.1038/ismej.2017.108.

Wang, L., Hatem, A., Catalyurek, U.V., Morrison, M., and Yu, Z. (2013). Metagenomic Insights into the Carbohydrate-Active Enzymes Carried by the Microorganisms Adhering to Solid Digesta in the Rumen of Cows. *PLoS ONE* 8(11)**,** 1-11. doi: 10.1371/journal.pone.0078507.

Weimer, P.J. (1998). Manipulating ruminal fermentation: a microbial ecological perspective. *Journal of Animal Science* 76(12), 3114-3122. doi: 10.2527/1998.76123114x.

Wolfger, B., Quinn, C., Torres, G.W., Taylor, M., and Orsel, K. (2016). Comparison of feeding behavior between black and red Angus feeder heifers. *Canadian Journal of Animal Science* 96(3), 404-409. doi: 10.1139/cjas-2014-0161.

Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15(3), R46-R46. doi: 10.1186/gb-2014-15-3-r46.

Zhou, M., Peng, Y.-J., Chen, Y., Klinger, C.M., Oba, M., Liu, J.-X., et al. (2018). Assessment of microbiome changes after rumen transfaunation: implications on improving feed efficiency in beef cattle. *Microbiome* 6(1), 62-62. doi: 10.1186/s40168-018-0447-y.

Bibliography

Aitchison, J. (1982) 'The statistical analysis of compositional data', *Journal of the Royal Statistical Society Series B-Methodological,* 44(2), 139-177.

Akin, D. E., Borneman, W. S. and Windham, W. R. (1988) 'RUMEN FUNGI - MORPHOLOGICAL TYPES FROM GEORGIA CATTLE AND THE ATTACK ON FORAGE CELL-WALLS', *Biosystems,* 21(3-4), 385-391.

Alemu, A. W., Vyas, D., Manafiazar, G., Basarab, J. A. and Beauchemin, K. A. (2017) 'Enteric methane emissions from low– and high–residual feed intake beef heifers measured using GreenFeed and respiration chamber techniques12', *Journal of Animal Science,* 95, 3727-3737.

Alexandratos, N. and Bruinsma, J. (2012) 'World Agriculture Towards 2030/2050: The 2012 Revision', *FAO - Food and Agriculture Organization of the United Nations*.

Anders, S. and Huber, W. (2010) 'Differential expression analysis for sequence count data', *Genome Biology,* 11(10), R106.

Anikaev, A., Isaev, A., Korobeinikova, A., Garber, M. and Gongadze, G. (2016) 'Role of protein L25 and its contact with protein L16 in maintaining the active state of Escherichia coli ribosomes in vivo', *Biochemistry (00062979),* 81(1), 19-27.

AOAC (2000) *Official Methods of Analysis (17th),* Gaithersburg, MD: Association of Official Analytical Chemists.

Archer, J. A., Richardson, E. C., Herd, R. M. and Arthur, P. F. (1999) 'Potential for selection to improve efficiency of feed use in beef cattle: a review', *Australian Journal of Agricultural Research,* 50(2), 147-162.

Artzi, L., Bayer, E. A. and Moraïs, S. (2016) 'Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides', *Nature Reviews Microbiology,* 15, 83.

Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. and Segata, N. (2015) 'Compact graphical representation of phylogenetic data and metadata with GraPhlAn', *PeerJ,* 3.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. and Zagnitko, O. (2008) 'The RAST Server: Rapid Annotations using Subsystems Technology', *BMC Genomics,* 9(1), 75.

Bader, J., Mast-Gerlach, E., Popovic, M. K., Bajpai, R. and Stahl, U. (2010) 'Relevance of microbial coculture fermentations in biotechnology', *Journal of Applied Microbiology,* 109(2), 371-387.

Bae, H. D., McAllister, T. A., Kokko, E. G., Leggett, F. L., Yanke, L. J., Jakober, K. D., Ha, J. K., Shin, H. T. and Cheng, K. J. (1997) 'Effect of silica on the colonization of rice straw by ruminal bacteria', *Animal Feed Science and Technology,* 65(1), 165-181.

Ban, Y., Jiang, H. and An, L. (2015) '- Investigating microbial co-occurrence patterns based on metagenomic compositional data', - 31(- 20), - 3329.

Bashiardes, S., Godneva, A., Elinav, E. and Segal, E. (2018) 'Towards utilization of the human genome and microbiome for personalized nutrition', *Current Opinion in Biotechnology,* 51, 57-63.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015) 'Fitting Linear Mixed-Effects Models Using lme4', *2015,* 67(1), 48.

Bayer, E. A., Chanzy, H., Lamed, R. and Shoham, Y. (1998) 'Cellulose, cellulases and cellulosomes', *Current Opinion in Structural Biology,* 8(5), 548-557.

Bayley, D. P., Correia, J. D., Jarrell, K. F. and Thomas, N. A. (1999) 'Recent Excitement about the Archaea: The Archaea are valuable for studying basic biological questions and have novel biotechnology applications', *BioScience,* 49(7), 530-541.

Beg, Q., Kapoor, M., Mahajan, L. and Hoondal, G. (2001) 'Microbial xylanases and their industrial applications: a review', *Applied Microbiology and Biotechnology,* 56(3), 326-338.

Belanche, A., Abecia, L., Holtrop, G., Guada, J. A., Castrillo, C., de la Fuente, G. and Balcells, J. (2011) 'Study of the effect of presence or absence of protozoa on rumen fermentation and microbial protein contribution to the chyme1', *Journal of Animal Science,* 89(12), 4163-4174.

Belanche, A., Doreau, M., Edwards, J. E., Moorby, J. M., Pinloche, E. and Newbold, C. J. (2012) 'Shifts in the Rumen Microbiota Due to the Type of Carbohydrate and Level of Protein Ingested by Dairy Cattle Are Associated with Changes in Rumen Fermentation', *Journal of Nutrition,* 142(9), 1684-1692.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate - a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society Series B-Methodological,* 57(1), 289-300.

Benndorf, D., Vogt, C., Jehmlich, N., Schmidt, Y., Thomas, H., Woffendin, G., Shevchenko, A., Richnow, H.-H. and von Bergen, M. (2009) 'Improving protein extraction and separation methods for investigating the metaproteome of

anaerobic benzene communities within sediments', *Biodegradation,* 20(6), 737-750.

Béra-Maillet, C., Mosoni, P., Kwasiborski, A., Suau, F., Ribot, Y. and Forano, E. (2009) 'Development of a RT-qPCR method for the quantification of Fibrobacter succinogenes S85 glycoside hydrolase transcripts in the rumen content of gnotobiotic and conventional sheep', *Journal Of Microbiological Methods,* 77(1), 8-16.

Berry, D. P. and Crowley, J. J. (2013) 'CELL BIOLOGY SYMPOSIUM: Genetics of feed efficiency in dairy and beef cattle1', *Journal of Animal Science,* 91(4), 1594-1613.

Bhat, M. K. and Bhat, S. (1997) 'Cellulose degrading enzymes and their potential industrial applications', *Biotechnology Advances,* 15(3), 583-620.

Bhatia, L., Johri, S. and Ahmad, R. (2012) 'An economic and ecological perspective of ethanol production from renewable agro waste: a review', *Amb Express,* 2.

Blazewicz, S. J., Barnard, R. L., Daly, R. A. and Firestone, M. K. (2013) 'Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses', *The ISME Journal,* 7, 2061.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics (Oxford, England),* 30(15), 2114-2120.

Borneman, W. S., Hartley, R. D., Morrison, W. H., Akin, D. E. and Ljungdahl, L. G. (1990) 'Feruloyl and p-coumaroyl esterase from anaerobic fungi in relation to plant cell wall degradation', *Applied Microbiology and Biotechnology,* 33(3), 345-351.

Brooker, J. D., O'Donovan, L. A., Skene, I., Clarke, K., Blackall, L. and Muslera, P. (1994) 'Streptococcus caprinus sp.nov., a tannin-resistant ruminal bacterium from feral goats', *Letters in Applied Microbiology,* 18(6), 313-318.

Brulc, J. M., Antonopoulos, D. A., Berg Miller, M. E., Wilson, M. K., Yannarell, A. C., Dinsdale, E. A., Edwards, R. E., Frank, E. D., Emerson, J. B., Wacklin, P., Coutinho, P. M., Henrissat, B., Nelson, K. E. and White, B. A. (2009) 'Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases', *Proceedings of the National Academy of Sciences,* 106(6), 1948-1953.

Bryant, M. P. and Burkey, L. A. (1953) 'Cultural Methods and Some Characteristics of Some of the More Numerous Groups of Bacteria in the Bovine Rumen', *Journal of Dairy Science,* 36(3), 205-217.

Bryant, M. P. and Robinson, I. M. (1961) 'An Improved Nonselective Culture Medium for Ruminal Bacteria and Its Use in Determining Diurnal Variation in Numbers of Bacteria in the Rumen', *Journal of Dairy Science,* 44(8), 1446-1456.

Bullard, J. H., Purdom, E., Hansen, K. D. and Dudoit, S. (2010) 'Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments', *Bmc Bioinformatics,* 11(1), 94.

Burnham, K. P., Anderson, D. R. and Burnham, K. P. (2002) *Model selection and multimodel inference : a practical information-theoretic approach,* New York : Springer, c2002.
2nd ed.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. and Holmes, S. P. (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods,* 13, 581.

Cammack, K. M., Austin, K. J., Lamberson, W. R., Conant, G. C. and Cunningham, H. C. (2018) 'Tiny but mighty: The role of the rumen microbes in livestock production', *J Anim Sci*.

Cantalapiedra-Hijar, G., Abo-Ismail, M., Carstens, G. E., Guan, L. L., Hegarty, R., Kenny, D. A., McGee, M., Plastow, G., Relling, A. and Ortigues-Marty, I. (2018) 'Review: Biological determinants of between-animal variation in feed efficiency of growing beef cattle', *Animal,* 12(s2), s321-s335.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) 'The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics', *Nucleic Acids Research,* 37, D233-D238.

Cao, K. A. L., Costello, M. E., Lakis, V. A., Bartolo, F., Chua, X. Y., Brazeilles, R. and Rondeau, P. (2016) 'MixMC: A multivariate statistical framework to gain insight into microbial communities', *Plos One,* 11(8).

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010) 'QIIME allows analysis of high-throughput community sequencing data', *Nature Methods,* 7(5), 335-336.

Carberry, C. A., Kenny, D. A., Han, S., McCabe, M. S. and Waters, S. M. (2012a) 'Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle', *Appl Environ Microbiol,* 78(14), 4949-58.

Carberry, C. A., Kenny, D. A., Kelly, A. K. and Waters, S. M. (2014a) 'Quantitative analysis of ruminal methanogenic microbial populations in beef cattle divergent in phenotypic residual feed intake (RFI) offered contrasting diets', *J Anim Sci Biotechnol,* 5(1), 41.

Carberry, C. A., Kenny, D. A., Sukkyan, H., McCabe, M. S. and Waters, S. M. (2012b) 'Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle', *Applied and Environmental Microbiology,* 78(14), 4949-4958.

Carberry, C. A., Waters, S. M., Kenny, D. A. and Creevey, C. J. (2014b) 'Rumen methanogenic genotypes differ in abundance according to host residual feed intake phenotype and diet type', *Appl Environ Microbiol,* 80(2), 586-94.

CCAC (1993) 'Guide to the care and use of experimental steers' in *Canadian Council on Animal Care*, 2nd ed. ed., Ottawa: Ottawa, Ontario, Canada.

Chafe, S. C. (1969) 'The fine structure of the collenchyma cell wall', *Planta,* 90(1), 12-21.

Cheng, K. J., McCowan, R. P. and Costerton, J. W. (1979) 'Adherent epithelial bacteria in ruminants and their roles in digestive tract function', *Am J Clin Nutr,* 32(1), 139-48.

Chomczynski, P. and Sacchi, N. (2006) 'The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on', *Nature Protocols,* (2), 581.

Christopherson, M. R., Dawson, J. A., Stevenson, D. M., Cunningham, A. C., Bramhacharya, S., Weimer, P. J., Kendziorski, C. and Suen, G. (2014) 'Unique aspects of fiber degradation by the ruminal ethanologen Ruminococcus albus 7 revealed by physiological and transcriptomic analysis', *BMC Genomics,* 15.

Ciucci, S., Ge, Y., Durán, C., Palladini, A., Martínez-Sánchez, L. M., Cannistraci, C. V., Gerl, M. J., Jiménez-Jiménez, V., Wang, Y., Sales, S., Shevchenko, A., Androutsellis-Theotokis, A., Poser, S. W., Herbig, M., Otto, O. and Guck, J. (2017) 'Enlightening discriminative network functional modules behind principal component analysis separation in differential-omic science studies', *Scientific Reports,* 7.

Clark, J. H., Budarin, V., Deswarte, F. E. I., Hardy, J. J. E., Kerton, F. M., Hunt, A. J., Luque, R., Macquarrie, D. J., Milkowski, K., Rodriguez, A., Samuel, O., Tavener, S. J., White, R. J. and Wilson, A. J. (2006) 'Green chemistry and the biorefinery: a partnership for a sustainable future', *Green Chemistry,* 8(10), 853-860.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R. and Tiedje, J. M. (2014) 'Ribosomal Database Project: data

and tools for high throughput rRNA analysis', *Nucleic Acids Research,* 42(D1), D633-D642.

Comtet-Marre, S., Chaucheyras-Durand, F., Bouzid, O., Mosoni, P., Bayat, A. R., Peyret, P. and Forano, E. (2018) 'FibroChip, a Functional DNA Microarray to Monitor Cellulolytic and Hemicellulolytic Activities of Rumen Microbiota', *Frontiers in Microbiology,* 9(215).

Comtet-Marre, S., Parisot, N., Lepercq, P., Chaucheyras-Durand, F., Mosoni, P., Peyretaillade, E., Bayat, A. R., Shingfield, K. J., Peyret, P. and Forano, E. (2017) 'Metatranscriptomics Reveals the Active Bacterial and Eukaryotic Fibrolytic Communities in the Rumen of Dairy Cow Fed a Mixed Diet', *Frontiers in Microbiology,* 8(67).

Cook, G. M., Rainey, F. A., Chen, G., Stackebrandt, E. and Russell, J. B. (1994) 'Emendation of the description of Acidaminococcus fermentans, a trans- aconitate- and citrate-oxidizing bacterium', *International Journal Of Systematic Bacteriology,* 44(3), 576-578.

Cox, M. P., Peterson, D. A. and Biggs, P. J. (2010) 'SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data', *Bmc Bioinformatics,* 11(1), 485.

Creevey, C. J., Kelly, W. J., Henderson, G. and Leahy, S. C. (2014) 'Determining the culturability of the rumen bacterial microbiome', *Microbial Biotechnology,* 7(5), 467-479.

Crews, D. H., Jr., Shannon, N. H., Genswein, B. M. A., Crews, R. E., Johnson, C. M. and Kendrick, B. A (2005) 'Genetic parameters for net feed efficiency of beef cattle measured during postweaning growing versus finishing periods', *Proc. West Sec. Am. Soc. Anim. Sci,* 54(54), 125-128.

Dai, X., Tian, Y., Li, J., Su, X., Wang, X., Zhao, S., Liu, L., Luo, Y., Liu, D., Zheng, H., Wang, J., Dong, Z., Hu, S. and Huang, L. (2015) 'Metatranscriptomic Analyses of Plant Cell Wall Polysaccharide Degradation by Microorganisms in the Cow Rumen', *Applied and Environmental Microbiology,* 81(4), 1375.

Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A. I. V., Bik, H. M. and Eisen, J. A. (2014) 'PhyloSift: phylogenetic analysis of genomes and metagenomes', *PeerJ,* 2, e243.

Dassa, B., Borovok, I., Ruimy-Israeli, V., Lamed, R., Flint, H. J., Duncan, S. H., Henrissat, B., Coutinho, P., Morrison, M., Mosoni, P., Yeoman, C. J., White, B. A. and Bayer, E. A. (2014) 'Rumen Cellulosomics: Divergent Fiber-Degrading Strategies Revealed by Comparative Genome-Wide Analysis of Six Ruminococcal Strains', *Plos One,* 9(7), e99221.

Davids, M., Hugenholtz, F., Martins dos Santos, V., Smidt, H., Kleerebezem, M. and Schaap, P. J. (2016) 'Functional Profiling of Unfamiliar Microbial Communities Using a Validated De Novo Assembly Metatranscriptome Pipeline', *Plos One,* 11(1), e0146423.

Davies, G. and Henrissat, B. (1995) 'Structures and mechanisms of glycosyl hydrolases', *Structure,* 3(9), 853-859.

Denman, S. E., Martinez Fernandez, G., Shinkai, T., Mitsumori, M. and McSweeney, C. S. (2015) 'Metagenomic analysis of the rumen microbial community following inhibition of methane formation by a halogenated methane analog', *Frontiers in Microbiology,* 6(1087).

Denman, S. E. and McSweeney, C. S. (2006) 'Development of a real-time PCR assay for monitoring anaerobic fungal and cellulolytic bacterial populations within the rumen', *Fems Microbiology Ecology,* 58(3), 572-582.

Denman, S. E., Morgavi, D. P. and McSweeney, C. S. (2018) 'Review: The application of omics to rumen microbiota function', *Animal,* 12(s2), s233-s245.

Deppenmeier, U. (2002) 'The unique biochemistry of methanogenesis', *Prog Nucleic Acid Res Mol Biol,* 71, 223-83.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G. L. (2006) 'Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB', *Applied and Environmental Microbiology,* 72(7), 5069-5072.

Deusch, S., Camarinha-Silva, A., Conrad, J., Beifuss, U., Rodehutscord, M. and Seifert, J. (2017) 'A Structural and Functional Elucidation of the Rumen Microbiome Influenced by Various Diets and Microenvironments', *Frontiers in Microbiology,* 8(1605).

Deusch, S., Tilocca, B., Camarinha-Silva, A. and Seifert, J. (2015) 'News in livestock research — use of Omics-technologies to study the microbiota in the gastrointestinal tract of farm animals', *Computational and Structural Biotechnology Journal,* 13, 55-63.

Distaso, M. A., Tran, H., Ferrer, M. and Golyshin, P. N. (2017) 'Metagenomic Mining of Enzyme Diversity' in Lee, S. Y., ed. *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals*, Cham: Springer International Publishing, 245-269.

Doi, R. H. and Kosugi, A. (2004) 'Cellulosomes: plant-cell-wall-degrading enzyme complexes', *Nature Reviews Microbiology,* 2, 541.

Dollhofer, V., Podmirseg, S. M., Callaghan, T. M., Griffith, G. W. and Fliegerova, K. (2015) 'Anaerobic Fungi and Their Potential for Biogas Production', *Advances in biochemical engineering/biotechnology,* 151, 41-61.

Duffield, T., Plaizier, J. C., Fairfield, A., Bagg, R., Vessie, G., Dick, P., Wilson, J., Aramini, J. and McBride, B. (2004) 'Comparison of techniques for measurement of rumen pH in lactating dairy cows', *Journal of Dairy Science,* 87(1), 59-66.

Ebringerová, A. and Heinze, T. (2000) 'Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties', *Macromolecular Rapid Communications,* 21(9), 542-556.

El Enshasy, H. A., Kandiyil, S. K., Malek, R. and Othman, N. Z. (2016) 'Microbial Xylanases: Sources, Types, and Their Applications' in Gupta, V. K., ed. *Microbial Enzymes in Bioconversions of Biomass*, Cham: Springer International Publishing, 151-213.

Fan, J., Han, F. and Liu, H. (2014) 'Challenges of Big Data analysis', *National Science Review,* 1(2), 293-314.

Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G. and Gloor, G. B. (2013) 'ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq', *Plos One,* 8(7), e67019.

Fernandes, A. D., Reid, J. N. S., Macklaim, J. M., McMurrough, T. A., Edgell, D. R. and Gloor, G. B. (2014) 'Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis', *Microbiome,* 2(1), 15.

Fernández-Arrojo, L., Guazzaroni, M.-E., López-Cortés, N., Beloqui, A. and Ferrer, M. (2010) 'Metagenomic era for biocatalyst identification', *Current Opinion in Biotechnology,* 21(6), 725-733.

Fernando, S. C., Purvis, H. T., II, Najar, F. Z., Sukharnikov, L. O., Krehbiel, C. R., Nagaraja, T. G., Roe, B. A. and DeSilva, U. (2010) 'Rumen Microbial Population Dynamics during Adaptation to a High-Grain Diet', *Applied and Environmental Microbiology,* 76(22), 7482-7490.

Ferrer, M., Beloqui, A., Timmis, K. N. and Golyshin, P. N. (2009) 'Metagenomics for Mining New Genetic Resources of Microbial Communities', *Journal of Molecular Microbiology and Biotechnology,* 16(1-2), 109-123.

Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Reyes-Duarte, D., Santos, V. A. P. M. D., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M. M., Timmis, K. N. and Golyshin, P. N. (2005) 'Novel hydrolase diversity retrieved

from a metagenome library of bovine rumen microflora', *Environmental Microbiology,* 7(12), 1996-2010.

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V. and Golyshin, P. N. (2016) 'Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends', *Microbial Biotechnology,* 9(1), 22-34.

Finlay, B. J., Esteban, G., Clarke, K. J., Williams, A. G., Embley, T. M. and Hirt, R. P. (1994) 'Some rumen ciliates have endosymbiotic methanogens', *Fems Microbiology Letters,* 117(2), 157-161.

Firkins, J. L. and Yu, Z. (2015) 'RUMINANT NUTRITION SYMPOSIUM: How to use data on the rumen microbiome to improve our understanding of ruminant nutrition', *J Anim Sci,* 93(4), 1450-70.

Ford, C. W., Elliott, R. and Maynard, P. J. (1987) 'THE EFFECT OF CHLORITE DELIGNIFICATION ON DIGESTIBILITY OF SOME GRASS FORAGES AND ON INTAKE AND RUMEN MICROBIAL ACTIVITY IN SHEEP FED BARLEY STRAW', *Journal of Agricultural Science,* 108, 129-136.

Foster, Z. S. L., Sharpton, T. J. and Grunwald, N. J. (2017) 'Metacoder: An R package for visualization and manipulation of community taxonomic diversity data', *Plos Computational Biology,* 13(2).

Fouts, D. E., Szpakowski, S., Purushe, J., Torralba, M., Waterman, R. C., MacNeil, M. D., Alexander, L. J. and Nelson, K. E. (2012) 'Next generation sequencing to define prokaryotic and fungal diversity in the bovine rumen', *PLoS One,* 7(11), e48289.

Franzosa, E. A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X. C. and Huttenhower, C. (2015) 'Sequencing and beyond: integrating molecular 'omics' for microbial community profiling', *Nature Reviews Microbiology,* 13, 360.

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., Giannoukos, G., Boylan, M. R., Ciulla, D., Gevers, D., Izard, J., Garrett, W. S., Chan, A. T. and Huttenhower, C. (2014) 'Relating the metatranscriptome and metagenome of the human gut', *Proceedings of the National Academy of Sciences,* 111(22), E2329-E2338.

Gao, A., Wang, H., Yang, J. and Shi, C. (2013) 'The effects of elimination of fungi on microbial population and fiber degradation in sheep rumen' in Tang, X., Zhong, W., Zhuang, D., Li, C. and Liu, Y., eds., *Progress in Environmental Protection and Processing of Resource, Pts 1-4*, 224-231.

Geishauser, T. (1993) 'An Instrument for collection and transfer of ruminal fluid and for administration of water soluble drugs in adult cattle', *The Bovine Practitioner,* (27), 38-42.

Gharechahi, J. and Salekdeh, G. H. (2018) 'A metagenomic analysis of the camel rumen's microbiome identifies the major microbes responsible for lignocellulose degradation and fermentation', *Biotechnology for Biofuels,* 11(1), 216.

Gijzen, H. J., Lubberding, H. J., Gerhardus, M. J. T. and Vogels, G. D. (1988) 'Contribution of rumen protozoa to fibre degradation and cellulase activity in vitro', *FEMS Microbiol Lett,* 53(1), 35-43.

Gloor, G., Wu, J. R., Pawlowsky-Glahn, V. and Jose Egozcue, J. (2016) '- It's all relative: analyzing microbiome data as compositions', - 26(- 5), - 329.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. and Egozcue, J. J. (2017) 'Microbiome Datasets Are Compositional: And This Is Not Optional', *Frontiers in Microbiology,* 8(2224).

Gloor, G. B. and Reid, G. (2016) 'Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data', *Canadian Journal of Microbiology,* 62(8), 692-703.

Goel, G., Puniya, A. K., Aguilar, C. N. and Singh, K. (2005) 'Interaction of gut microflora with tannins in feeds', *Naturwissenschaften,* 92(11), 497-503.

Gomez-Alvarez, V., Teal, T. K. and Schmidt, T. M. (2009) 'Systematic artifacts in metagenomes from complex microbial communities', *The ISME Journal,* 3, 1314.

Gonzalez, I., Dejean, S., Martin, P. G. P. and Baccini, A. (2008) 'CCA: An R package to extend canonical correlation analysis', *Journal of Statistical Software,* 23(12), 1-14.

Goodrich, J. K., Davenport, E. R., Clark, A. G. and Ley, R. E. (2017) 'The Relationship Between the Human Genome and Microbiome Comes into View', *Annual Review of Genetics,* 51(1), 413-433.

Gordon, G. L. R. and Phillips, M. W. (1992) 'Extracellular pectin lyase produced by Neocallimastix sp. LM1, a rumen anaerobic fungus', *Letters in Applied Microbiology,* 15(3), 113-115.

Gordon, G. L. R. and Phillips, M. W. (1993) 'REMOVAL OF ANAEROBIC FUNGI FROM THE RUMEN OF SHEEP BY CHEMICAL TREATMENT AND THE EFFECT ON FEED CONSUMPTION AND IN-VIVO FIBER DIGESTION', *Letters in Applied Microbiology,* 17(5), 220-223.

Gordon, G. L. R. and Phillips, M. W. (1998) 'The role of anaerobic gut fungi in ruminants', *Nutrition Research Reviews,* 11(1), 133-168.

Gottschalk, G. (1986) 'Bacterial metabolism',

Greve, L. C., Labavitch, J. M. and Hungate, R. E. (1984) 'alpha-L-arabinofuranosidase from Ruminococcus albus 8: purification and possible role in hydrolysis of alfalfa cell wall', *Applied and Environmental Microbiology,* 47(5), 1135-1140.

Grubb, J. A. and Dehority, B. A. (1975) 'EFFECTS OF AN ABRUPT CHANGE IN RATION FROM ALL ROUGHAGE TO HIGH CONCENTRATE UPON RUMEN MICROBIAL NUMBERS IN SHEEP', *Applied Microbiology,* 30(3), 404-412.

Gruninger, R. J., Puniya, A. K., Callaghan, T. M., Edwards, J. E., Youssef, N., Dagar, S. S., Fliegerova, K., Griffith, G. W., Forster, R., Tsang, A., McAllister, T. and Elshahed, M. S. (2014) 'Anaerobic fungi (phylum Neocallimastigomycota): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential', *Fems Microbiology Ecology,* 90(1), 1-17.

Guan, L. L., Nkrumah, J. D., Basarab, J. A. and Moore, S. S. (2008) 'Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency', *Fems Microbiology Letters,* 288(1), 85-91.

Gunsett, F. C. (1984) 'Linear Index Selection to Improve Traits Defined as Ratios', *Journal of Animal Science,* 59(5), 1185-1193.

Gupta, A., Kumar, S., Prasoodanan, V. P. K., Harish, K., Sharma, A. K. and Sharma, V. K. (2016a) 'Reconstruction of Bacterial and Viral Genomes from Multiple Metagenomes', *Front Microbiol,* 7, 469.

Gupta, A., Kumar, S., Prasoodanan, V. P. K., Harish, K., Sharma, A. K. and Sharma, V. K. (2016b) 'Reconstruction of Bacterial and Viral Genomes from Multiple Metagenomes', *Frontiers in Microbiology,* 7(469).

Hackmann, T. J. and Firkins, J. L. (2015) 'Maximizing efficiency of rumen microbial protein production', *Frontiers in Microbiology,* 6.

Hackmann, T. J., Tao, J., Ngugi, D. K. and Firkins, J. L. (2017) 'Genomes of rumen bacteria encode atypical pathways for fermenting hexoses to short-chain fatty acids', *Environmental Microbiology,* 19(11), 4670-4683.

Hall, M. B. (2009) 'Determination of Starch, Including Maltooligosaccharides, in Animal Feeds: Comparison of Methods and a Method Recommended for AOAC Collaborative Study', *Journal of Aoac International,* 92(1), 42-49.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. and Goodman, R. M. (1998) 'Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products', *Chemistry & Biology,* 5(10), R245-R249.

Hart, E. H., Creevey, C. J., Hitch, T. and Kingston-Smith, A. H. (2018) 'Meta-proteomics of rumen microbiota indicates niche compartmentalisation and functional dominance in a limited number of metabolic pathways between abundant bacteria', *Scientific Reports,* 8(1), 10504.

Healey, G. R., Murphy, R., Brough, L., Butts, C. A. and Coad, J. (2017) 'Interindividual variability in gut microbiota and host response to dietary interventions', *Nutrition Reviews,* 75(12), 1059-1080.

Hegarty, R. S., Goopy, J. P., Herd, R. M. and McCorkell, B. (2007) 'Cattle selected for lower residual feed intake have reduced daily methane production', *J Anim Sci,* 85(6), 1479-86.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W. and Janssen, P. H. (2015) 'Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range', *Sci Rep,* 5, 14567.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P. H. and Global Rumen Census, C. (2015) 'Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range', *Scientific Reports,* 6.

Hernandez-Sanabria, E., Goonewardene, L. A., Wang, Z., Durunna, O. N., Moore, S. S. and Guan, L. L. (2012) 'Impact of Feed Efficiency and Diet on Adaptive Variations in the Bacterial Community in the Rumen Fluid of Cattle', *Applied and Environmental Microbiology,* 78(4), 1203-1214.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D. S., Chen, F., Zhang, T., Mackie, R. I., Pennacchio, L. A., Tringe, S. G., Visel, A., Woyke, T., Wang, Z. and Rubin, E. M. (2011) 'Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen', *Science,* 331(6016), 463-467.

Himmel, M. E., Ding, S.-Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W. and Foust, T. D. (2007) 'Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production', *Science,* (5813).

Ho, Y. W., Abdullah, N. and Jalaludin, S. (1988) 'PENETRATING STRUCTURES OF ANAEROBIC RUMEN FUNGI IN CATTLE AND SWAMP BUFFALO', *Journal of General Microbiology,* 134(1), 177-182.

Ho, Y. W. and Barr, D. J. S. (1995) 'CLASSIFICATION OF ANAEROBIC GUT FUNGI FROM HERBIVORES WITH EMPHASIS ON RUMEN FUNGI FROM MALAYSIA', *Mycologia,* 87(5), 655-677.

Hobson, P. N. and Stewart, C. S. (1997) *The Rumen Microbial Ecosystem,* London: Blackie Academic & Professional.

Hogg, D., Pell, G., Dupree, P., Goubet, F., Martín-Orúe, S. M., Armand, S. and Gilbert, H. J. (2003) 'The modular architecture of Cellvibrio japonicus mannanases in glycoside hydrolase families 5 and 26 points to differences in their role in mannan degradation', *The Biochemical journal,* 371(Pt 3), 1027-1043.

Holtshausen, L., Schwartzkopf-Genswein, K. S. and Beauchemin, K. A. (2013) 'Short Communication: Ruminal pH profile and feeding behaviour of feedlot cattle transitioning from a high-forage to a high-concentrate diet', *Canadian Journal of Animal Science,* 93(4), 529-533.

Hristov, A. N., Callaway, T. R., Lee, C. and Dowd, S. E. (2012) 'Rumen bacterial, archaeal, and fungal diversity of dairy cows in response to ingestion of lauric or myristic acid', *J Anim Sci,* 90(12), 4449-57.

Hristov, A. N., Oh, J., Firkins, J. L., Dijkstra, J., Kebreab, E., Waghorn, G., Makkar, H. P. S., Adesogan, A. T., Yang, W., Lee, C., Gerber, P. J., Henderson, B. and Tricarico, J. M. (2013) 'SPECIAL TOPICS — Mitigation of methane and nitrous oxide emissions from animal operations: I. A review of enteric methane mitigation options', *Journal of Animal Science,* 91, 5045-5069.

Hristov, A. N., Oh, J., Giallongo, F., Frederick, T. W., Harper, M. T., Weeks, H. L., Branco, A. F., Moate, P. J., Deighton, M. H., Williams, S. R. O., Kindermann, M. and Duval, S. (2015) 'An inhibitor persistently decreased enteric methane emission from dairy cows with no negative effect on milk production', *Proceedings of the National Academy of Sciences,* 112(34), 10663-10668.

Hungate, R. E. (1966) *The rumen and its microbes,* New York and London: Academic Press.

Hungate, R. E. (1967) 'Hydrogen as an intermediate in the rumen fermentation', *Archiv für Mikrobiologie,* 59(1-3), 158-164.

Huws, S. A., Creevey, C. J., Oyama, L. B., Mizrahi, I., Denman, S. E., Popova, M., Muñoz-Tamayo, R., Forano, E., Waters, S. M., Hess, M., Tapio, I., Smidt, H., Krizsan, S. J., Yáñez-Ruiz, D. R., Belanche, A., Guan, L., Gruninger, R. J., McAllister, T. A., Newbold, C. J., Roehe, R., Dewhurst, R. J., Snelling, T. J., Watson, M., Suen, G., Hart, E. H., Kingston-Smith, A. H., Scollan, N. D., do Prado, R. M., Pilau, E. J., Mantovani, H. C., Attwood, G. T., Edwards, J. E., McEwan, N. R., Morrisson, S., Mayorga, O. L., Elliott, C. and Morgavi, D. P. (2018) 'Addressing Global Ruminant

Agricultural Challenges Through Understanding the Rumen Microbiome: Past, Present, and Future', *Frontiers in Microbiology,* 9(2161).

Imelfort, M., Parks, D., Woodcroft, B. J., Dennis, P., Hugenholtz, P. and Tyson, G. W. (2014) 'GroopM: an automated tool for the recovery of population genomes from related metagenomes', *PeerJ,* 2, e603.

IPCC (2006) *Guidelines for National Greenhouse Gas Inventories, Agriculture, forestry and other land use,* Japan: Institute for Global Environmental Strategies (IGES).

Jami, E., Israel, A., Kotser, A. and Mizrahi, I. (2013) 'Exploring the bovine rumen bacterial community from birth to adulthood', *Isme j,* 7(6), 1069-79.

Jami, E. and Mizrahi, I. (2012) 'Composition and similarity of bovine rumen microbiota across individual animals', *PLoS One,* 7(3), e33306.

Janssen, P. H. (2010) 'Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics', *Animal Feed Science and Technology,* 160(1-2), 1-22.

Janssen, P. H. and Kirs, M. (2008) 'Structure of the archaeal community of the rumen', *Appl Environ Microbiol,* 74(12), 3619-25.

Janssen, P. H. and Kirs, M. (2008) 'Structure of the Archaeal Community of the Rumen', *Applied & Environmental Microbiology,* 74(12), 3619-3625.

Jayanegara, A., Sarwono, K. A., Kondo, M., Matsui, H., Ridla, M., Laconi, E. B. and Nahrowi (2018) 'Use of 3-nitrooxypropanol as feed additive for mitigating enteric methane emissions from ruminants: a meta-analysis AU - Jayanegara, Anuraga', *Italian Journal of Animal Science,* 17(3), 650-656.

Jiménez, D. J., Montaña, J. S., Álvarez, D. and Baena, S. (2012) 'A novel cold active esterase derived from Colombian high Andean forest soil metagenome', *World Journal of Microbiology and Biotechnology,* 28(1), 361-370.

Jin, D., Zhao, S., Zheng, N., Bu, D., Beckers, Y., Denman, S. E., McSweeney, C. S. and Wang, J. (2017) 'Differences in Ureolytic Bacterial Composition between the Rumen Digesta and Rumen Wall Based on ureC Gene Classification', *Frontiers in Microbiology,* 8(385).

Johnson, D. E., Ferrell, C. L. and Jenkins, T. G. (2003) 'The history of energetic efficiency research: Where have we been and where are we going?', *Journal of Animal Science,* 81(13_suppl_1), E27-E38.

Johnson, K. A. and Johnson, D. E. (1995) 'Methane emissions from cattle', *Journal of Animal Science,* 73(8), 2483-2492.

Jones, D. R., Uddin, M. S., Gruninger, R. J., Pham, T. T. M., Thomas, D., Boraston, A. B., Briggs, J., Pluvinage, B., McAllister, T. A., Forster, R. J., Tsang, A., Selinger, L. B. and Abbott, D. W. (2017) 'Discovery and characterization of family 39 glycoside hydrolases from rumen anaerobic fungi with polyspecific activity on rare arabinosyl substrates', *Journal of Biological Chemistry,* 292(30), 12606-12620.

Joshi, M. D., Sidhu, G., Pot, I., Brayer, G. D., Withers, S. G. and McIntosh, L. P. (2000) 'Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the ph optimum of a glycosidase11Edited by M. F. Summers', *Journal Of Molecular Biology,* 299(1), 255-279.

Jouany, J. P. (1995) 'Rumen microbial metabolism and ruminant digestion: J-P. Jouany (Ed). 1991. INRA, 147 rue de l'Université, 75338 Paris cedex 07, France. ISBN 2 7380 0345 1. 374 pp. Text in English. 350 FF', *Livestock Production Science,* 44(3), 284.

Kabel, M. A., Yeoman, C. J., Han, Y., Dodd, D., Abbas, C. A., de Bont, J. A. M., Morrison, M., Cann, I. K. O. and Mackie, R. I. (2011) 'Biochemical Characterization and Relative Expression Levels of Multiple Carbohydrate Esterases of the Xylanolytic Rumen Bacterium', *Applied and Environmental Microbiology,* 77(16), 5671-5681.

Kaminski, J., Gibson, M. K., Franzosa, E. A., Segata, N., Dantas, G. and Huttenhower, C. (2015) 'High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED', *Plos Computational Biology,* 11(12).

Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., Janssen, P. H., Shi, W., Froula, J., Rubin, E. M. and Attwood, G. T. (2016) 'Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation', *Microbiome,* 4(1), 56-56.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) 'The KEGG resource for deciphering the genome', *Nucleic Acids Research,* 32(Database issue), D277-D280.

Kang, D. D., Froula, J., Egan, R. and Wang, Z. (2015) 'MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities', *PeerJ,* 3, e1165.

Kenters, N., Henderson, G., Jeyanathan, J., Kittelmann, S. and Janssen, P. H. (2011) 'Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium', *Journal Of Microbiological Methods,* 84(1), 52-60.

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

Kittelmann, S., Devente, S. R., Kirk, M. R., Seedorf, H., Dehority, B. A. and Janssen, P. H. (2015) 'Phylogeny of Intestinal Ciliates, Including Charonina ventriculi, and Comparison of Microscopy and 18S rRNA Gene Pyrosequencing for Rumen Ciliate Community Structure Analysis', *Applied and Environmental Microbiology,* 81(7), 2433.

Kittelmann, S., Devente, S. R., Kirk, M. R., Seedorf, H., Dehority, B. A. and Janssen, P. H. (2015) 'Phylogeny of Intestinal Ciliates, Including Charonina ventriculi, and Comparison of Microscopy and 18S rRNA Gene Pyrosequencing for Rumen Ciliate Community Structure Analysis', *Appl Environ Microbiol,* 81(7), 2433-2444.

Kittelmann, S., Pinares-Patiño, C. S., Seedorf, H., Kirk, M. R., Ganesh, S., McEwan, J. C. and Janssen, P. H. (2014) 'Two Different Bacterial Community Types Are Linked with the Low-Methane Emission Trait in Sheep', *Plos One,* 9(7), 1-9.

Kittelmann, S., Seedorf, H., Walters, W. A., Clemente, J. C., Knight, R., Gordon, J. I. and Janssen, P. H. (2013a) 'Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities', *PLoS One,* 8(2), e47879.

Kittelmann, S., Seedorf, H., Walters, W. A., Clemente, J. C., Knight, R., Gordon, J. I. and Janssen, P. H. (2013b) 'Simultaneous Amplicon Sequencing to Explore Co-Occurrence Patterns of Bacterial, Archaeal and Eukaryotic Microorganisms in Rumen Microbial Communities', *Plos One,* 8(2).

Knapp, J. R., Laur, G. L., Vadas, P. A., Weiss, W. P. and Tricarico, J. M. (2014) 'Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions', *J Dairy Sci,* 97(6), 3231-3261.

Knights, D., Parfrey, L. W., Zaneveld, J., Lozupone, C. and Knight, R. (2011) 'Human-associated microbial signatures: examining their predictive value', *Cell Host & Microbe,* 10(4), 292-296.

Koch, R. M., Swiger, L. A., Chambers, D. and Gregory, K. E. (1963) 'Efficiency of feed use in beef cattle', *Journal of Animal Science,* 22(2), 486-494.

Koetschan, C., Kittelmann, S., Lu, J., Al-Halbouni, D., Jarvis, G. N., Mueller, T., Wolf, M. and Janssen, P. H. (2014) 'Internal Transcribed Spacer 1 Secondary Structure Analysis Reveals a Common Core throughout the Anaerobic Fungi (Neocallimastigomycota)', *Plos One,* 9(3).

Kong, R. S., Liang, G., Chen, Y., Stothard, P. and Guan le, L. (2016) 'Transcriptome profiling of the rumen epithelium of beef cattle differing in residual feed intake', *BMC Genomics,* 17, 592.

Kopylova, E., Noé, L. and Touzet, H. (2012) 'SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data', *Bioinformatics (Oxford, England),* 28(24), 3211-3217.

Korver, S. (1988) 'Genetic aspects of feed intake and feed efficiency in dairy cattle: A review', *Livestock Production Science,* 20(1), 1-13.

Kosugi, A., Murashima, K. and Doi, R. H. (2001) 'Characterization of Xylanolytic Enzymes in Clostridium cellulovorans: Expression of Xylanase Activity Dependent on Growth Substrates', *Journal of Bacteriology,* 183(24), 7037-7043.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. (2013) 'Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform', *Applied and Environmental Microbiology,* 79(17), 5112-5120.

Krause, D. O., Nagaraja, T. G., Wright, A. D. G. and Callaway, T. R. (2013) 'Board-invited review: Rumen microbiology: Leading the way in microbial ecology1,2', *Journal of Animal Science,* 91(1), 331-341.

Kumar, S., Choudhury, P. K., Carro, M. D., Griffith, G. W., Dagar, S. S., Puniya, M., Calabro, S., Ravella, S. R., Dhewa, T., Upadhyay, R. C., Sirohi, S. K., Kundu, S. S., Wanapat, M. and Puniya, A. K. (2014) 'New aspects and strategies for methane mitigation from ruminants', *Appl Microbiol Biotechnol,* 98(1), 31-44.

Kurtz, Z. D., Mueller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J. and Bonneau, R. A. (2014) 'Sparse and compositionally robust inference of microbial ecological networks',

Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D. and Charles, T. C. (2015) 'Current and future resources for functional metagenomics', *Frontiers in Microbiology,* 6(1196).

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., Burkepile, D. E., Vega Thurber, R. L., Knight, R., Beiko, R. G. and Huttenhower, C. (2013) 'Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences', *Nature Biotechnology,* 31, 814.

Larance, M. and Lamond, A. I. (2015) 'Multidimensional proteomics for cell biology', *Nature Reviews Molecular Cell Biology,* 16, 269.

Larsbrink, J., Tuveng, T. R., Pope, P. B., Bulone, V., Eijsink, V. G. H., Brumer, H. and McKee, L. S. (2017a) 'Proteomic insights into mannan degradation and protein secretion by the forest floor bacterium Chitinophaga pinensis', *Journal of Proteomics,* 156, 63-74.

Larsbrink, J., Tuveng, T. R., Pope, P. B., Eijsink, V. G. H., Bulone, V., Brumer, H. and McKee, L. S. (2017b) 'Proteomic data on enzyme secretion and activity in the bacterium Chitinophaga pinensis', *Data in Brief,* 11, 484-490.

Leahy, S. C., Kelly, W. J., Altermann, E., Ronimus, R. S., Yeoman, C. J., Pacheco, D. M., Li, D., Kong, Z. H., McTavish, S., Sang, C., Lambie, S. C., Janssen, P. H., Dey, D. and Attwood, G. T. (2010) 'The Genome Sequence of the Rumen Methanogen Methanobrevibacter ruminantium Reveals New Possibilities for Controlling Ruminant Methane Emissions', *Plos One,* 5(1).

Lee, S., Sung, J., Lee, J. and Ko, G. (2011) 'Comparison of the gut microbiotas of healthy adult twins living in South Korea and the United States', *Applied and Environmental Microbiology,* 77(20), 7433-7437.

Lee, S. S., Ha, J. K. and Cheng, K. J. (2000) 'Relative contributions of bacteria, protozoa, and fungi to in vitro degradation of orchard grass cell walls and their interactions', *Applied and Environmental Microbiology,* 66(9), 3807-3813.

Levin, B. J., Huang, Y. Y., Peck, S. C., Wei, Y., Campo, A. M. D., Marks, J. A., Franzosa, E. A., Huttenhower, C. and Balskus, E. P. (2017) 'A prominent glycyl radical enzyme in human gut microbiomes metabolizes trans-4-hydroxy-L-proline', *Science,* 355(6325), 595-+.

Li, D. H., Liu, C. M., Luo, R. B., Sadakane, K. and Lam, T. W. (2015) 'MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics,* 31(10), 1674-1676.

Li, F. and Guan, L. L. (2017) 'Metatranscriptomic Profiling Reveals Linkages between the Active Rumen Microbiome and Feed Efficiency in Beef Cattle', *Appl Environ Microbiol,* 83(9).

Li, F. and Guan, L. L. (2017) 'Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle', *Applied and Environmental Microbiology*.

Li, F., Henderson, G., Sun, X., Cox, F., Janssen, P. H. and Guan le, L. (2016) 'Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches', *Front Microbiol,* 7, 987.

Li, F., Henderson, G., Sun, X., Cox, F., Janssen, P. H. and Guan, L. L. (2016) 'Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches', *Frontiers in Microbiology,* 7(987).

Li, F., Hitch, T. C. A., Chen, Y., Creevey, C. J. and Guan, L. L. (2019a) 'Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the

rumen microbiome and its associations with feed efficiency in beef cattle', *Microbiome,* 7(1), 6.

Li, F., Li, C., Chen, Y., Liu, J., Zhang, C., Irving, B., Fitzsimmons, C., Plastow, G. and Guan, L. L. (2019b) 'Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle', *Microbiome,* 7(1), 92.

Li, M., Penner, G. B., Hernandez-Sanabria, E., Oba, M. and Guan, L. L. (2009) 'Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen', *Journal of Applied Microbiology,* 107(6), 1924-1934.

Li, R. W., Connor, E. E., Li, C., Baldwin Vi, R. L. and Sparks, M. E. (2012) 'Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools', *Environ Microbiol,* 14(1), 129-39.

Li, Y. and Zhang, Y. (2009) 'REMO: A New Protocol to Refine Full Atomic Protein Models from C-alpha Traces by Optimizing Hydrogen-Bonding Networks', *Proteins,* 76(3), 665-676.

Li, Z. P., Wright, A. D. G., Si, H. Z., Wang, X. X., Qian, W. X., Zhang, Z. G. and Li, G. Y. (2016) 'Changes in the rumen microbiome and metabolites reveal the effect of host genetics on hybrid crosses', *Environmental Microbiology Reports,* 8(6), 1016-1023.

Lin, C., Stahl, D. A. and Raskin, L. (1997) 'Microbial community structure in gastrointestinal tracts of domestic animals: comparative analyses using rRNA-targeted oligonucleotide probes', *Fems Microbiology Ecology,* 22(4), 281-294.

Liquet, B., Cao, K.-A. L., Hocini, H. and Thiébaut, R. (2012) 'A novel approach for biomarker selection and the integration of repeated measures experiments from two assays', *Bmc Bioinformatics,* 13(1), 325.

Liu, J. h., Zhang, M. l., Zhang, R. y., Zhu, W. y. and Mao, S. y. (2016) 'Comparative studies of the composition of bacterial microbiota associated with the ruminal content, ruminal epithelium and in the faeces of lactating dairy cows', *Microb Biotechnol,* 9(2), 257-268.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B. (2014) 'The carbohydrate-active enzymes database (CAZy) in 2013', *Nucleic Acids Research,* 42(D1), D490-D495.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology,* 15(12), 550-550.

Lovell, D., Pawlowsky-Glahn, V., Egozcue, J. J., Marguerat, S. and Bähler, J. (2015) 'Proportionality: A Valid Alternative to Correlation for Relative Data', *Plos Computational Biology,* 11(3), 1-12.

Malmuthuge, N., Li, M., Chen, Y., Fries, P., Griebel, P. J., Baurhoo, B., Zhao, X. and Guan, L. L. (2012) 'Distinct commensal bacteria associated with ingesta and mucosal epithelium in the gastrointestinal tracts of calves and chickens', *Fems Microbiology Ecology,* 79(2), 337-347.

Mandal, S., Van Treuren, W., White, R. A., Eggesbo, M., Knight, R. and Peddada, S. D. (2015) 'Analysis of composition of microbiomes: a novel method for studying microbial composition', *Microbial Ecology in Health and Disease,* 26, 27663-Article No.: 27663.

Marchesi, J. R. and Ravel, J. (2015) 'The vocabulary of microbiome research: a proposal', *Microbiome,* 3(1), 31.

Márquez, G. C., Speidel, S. E., Enns, R. M. and Garrick, D. J. (2010) 'Genetic diversity and population structure of American Red Angus cattle', *Journal of Animal Science,* 88(1), 59-68.

Martin, C., Morgavi, D. P. and Doreau, M. (2010) 'Methane mitigation in ruminants: from microbe to the farm scale', *animal,* 4(3), 351-65.

Martínez, I., Jaehyoung, K., Duffy, P. R., Schlegel, V. L. and Walter, J. (2010) 'Resistant Starches Types 2 and 4 Have Differential Effects on the Composition of the Fecal Microbiota in Human Subjects', *Plos One,* 5(11), 1-11.

Martinez-Fernandez, G., Denman, S. E., Cheung, J. and McSweeney, C. S. (2017) 'Phloroglucinol Degradation in the Rumen Promotes the Capture of Excess Hydrogen Generated from Methanogenesis Inhibition', *Frontiers in Microbiology,* 8(1871).

Martinez-Fernandez, G., Denman, S. E., Yang, C., Cheung, J., Mitsumori, M. and McSweeney, C. S. (2016) 'Methane Inhibition Alters the Microbial Community, Hydrogen Flow, and Fermentation Response in the Rumen of Cattle', *Frontiers in Microbiology,* 7(1122).

Matthews, C., Crispie, F., Lewis, E., Reid, M., O'Toole, P. W. and Cotter, P. D. (2018) 'The rumen microbiome: a crucial consideration when optimising milk and meat production and nitrogen utilisation efficiency AU - Matthews, Chloe', *Gut Microbes*, 1-18.

McAllister, T. A., Bae, H. D., Jones, G. A. and Cheng, K. J. (1994a) 'Microbial attachment and feed digestion in the rumen', *Journal of Animal Science,* 72(11), 3004-3018.

McAllister, T. A., Bae, H. D., Yanke, L. J., Cheng, K. J. and Muir, A. (1994b) 'Effect of condensed tannins from birdsfoot trefoil on endoglucanase activity and the digestion of cellulose filter paper by ruminal fungi', *Can J Microbiol,* 40(4), 298-305.

McAllister, T. A., Meale, S. J., Valle, E., Guan, L. L., Zhou, M., Kelly, W. J., Henderson, G., Attwood, G. T. and Janssen, P. H. (2015) 'RUMINANT NUTRITION SYMPOSIUM: Use of genomics and transcriptomics to identify strategies to lower ruminal methanogenesis', *J Anim Sci,* 93(4), 1431-49.

McCabe, M. S., Cormican, P., Keogh, K., O'Connor, A., O'Hara, E., Palladino, R. A., Kenny, D. A. and Waters, S. M. (2015) 'Illumina MiSeq Phylogenetic Amplicon Sequencing Shows a Large Reduction of an Uncharacterised Succinivibrionaceae and an Increase of the Methanobrevibacter gottschalkii Clade in Feed Restricted Cattle', *PLoS One,* 10(7), e0133234.

McCabe, M. S., Cormican, P., Keogh, K., O'Connor, A., O'Hara, E., Palladino, R. A., Kenny, D. A. and Waters, S. M. (2015) 'Illumina MiSeq Phylogenetic Amplicon Sequencing Shows a Large Reduction of an Uncharacterised Succinivibrionaceae and an Increase of the Methanobrevibacter gottschalkii Clade in Feed Restricted Cattle', *Plos One,* 10(7), e0133234.

McCann, J. C., Elolimy, A. A. and Loor, J. J. (2017) 'Rumen Microbiome, Probiotics, and Fermentation Additives', *The Veterinary Clinics Of North America. Food Animal Practice*.

McCarthy, D. J., Chen, Y. and Smyth, G. K. (2012) 'Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation', *Nucleic Acids Research,* 40(10), 4288-4297.

McDougall, E. I. (1948) 'Studies on ruminant saliva. 1. The composition and output of sheep's saliva', *Biochemical Journal,* 43(1), 99-109.

McGovern, E., McCabe, M. S., Cormican, P., Popova, M., Keogh, K., Kelly, A. K., Kenny, D. A. and Waters, S. M. (2017) 'Plane of nutrition affects the phylogenetic diversity and relative abundance of transcriptionally active methanogens in the bovine rumen', *Scientific Reports,* 7(1), 13047.

McLean, K. L. and Schmutz, S. M. (2009) 'Associations of melanocortin 1 receptor genotype with growth and carcass traits in beef cattle', *Canadian Journal of Animal Science,* 89(3), 295-300.

McMurdie, P. J. and Holmes, S. (2014) 'Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible', *Plos Computational Biology,* 10(4), e1003531.

McNeil, M., Darvill, A. G., Fry, S. C. and Albersheim, P. (1984) 'Structure and Function of the Primary Cell Walls of Plants', *Annual Review of Biochemistry,* 53(1), 625-663.

McSweeney, C. and Mackie, R. I. (2012) *Microorganisms and ruminant digestion: state of knowledge, trends and future prospects, Comission on genetic resources for food and agriculture*

Mesuere, B., Devreese, B., Debyser, G., Aerts, M., Vandamme, P. and Dawyndt, P. (2012) 'Unipept: Tryptic Peptide-Based Biodiversity Analysis of Metaproteome Samples', *Journal of Proteome Research,* 11(12), 5773-5780.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J. and Edwards, R. (2008a) 'The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes', *Bmc Bioinformatics,* 9(1), 386.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J. and Edwards, R. A. (2008b) 'The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes', *Bmc Bioinformatics,* 9.

Miller, G. L. (1959) 'Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar', *Analytical Chemistry,* 31(3), 426-428.

Miron, J., Ben-Ghedalia, D. and Morrison, M. (2001) 'Invited Review: Adhesion Mechanisms of Rumen Cellulolytic Bacteria', *Journal of Dairy Science,* 84(6), 1294-1309.

Montanholi, Y. R., Swanson, K. C., Palme, R., Schenkel, F. S., McBride, B. W., Lu, D. and Miller, S. P. (2010) 'Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids', *Animal,* 4(5), 692-701.

Morgavi, D., Kelly, W. J., Janssen, P. H. and Attwood, G. T. (2013) 'Rumen microbial (meta) genomics and its application to ruminant production', 184.

Morgavi, D. P., Forano, E., Martin, C. and Newbold, C. J. (2010) 'Microbial ecosystem and methanogenesis in ruminants', *Animal,* 4(7), 1024-1036.

Morgavi, D. P., Kelly, W. J., Janssen, P. H. and Attwood, G. T. (2012a) 'Rumen microbial (meta)genomics and its application to ruminant production', *Animal,* 7(s1), 184-201.

Morgavi, D. P., Martin, C., Jouany, J. P. and Ranilla, M. J. (2012b) 'Rumen protozoa and methanogenesis: not a simple cause-effect relationship', *Br J Nutr,* 107(3), 388-97.

Morrison, M., Murray, R. M. and Boniface, A. N. (1990) 'NUTRIENT METABOLISM AND RUMEN MICROORGANISMS IN SHEEP FED A POOR-QUALITY TROPICAL GRASS HAY SUPPLEMENTED WITH SULFATE', *Journal of Agricultural Science,* 115, 269-275.

Munoz-Munoz, J., Cartmell, A., Terrapon, N., Henrissat, B. and Gilbert, H. J. (2017) 'Unusual active site location and catalytic apparatus in a glycoside hydrolase family', *Proceedings of the National Academy of Sciences*.

Myer, P. R., Smith, T. P. L., Wells, J. E., Kuehn, L. A. and Freetly, H. C. (2015a) 'Rumen Microbiome from Steers Differing in Feed Efficiency', *PLoS One,* 10(6), e0129174.

Myer, P. R., Smith, T. P. L., Wells, J. E., Kuehn, L. A. and Freetly, H. C. (2015b) 'Rumen Microbiome from Steers Differing in Feed Efficiency', *Plos One,* 10(6).

Nations, U. (2015) 'World Population Prospects', [online], available].

Nayfach, S., Bradley, P. H., Wyman, S. K., Laurent, T. J., Williams, A., Pollard, K. S., Sharpton, T. J. and Eisen, J. A. (2015) 'Automated and accurate estimation of gene family abundance from shotgun metagenomes', *Plos Computational Biology,* 11(11), 29p.

Network, I. F. C. (2016) *Long-term dairy outlook*.

Neves, A. L. A., Li, F., Ghoshal, B., McAllister, T. and Guan, L. L. (2017) 'Enhancing the resolution of rumen microbial classification from metatranscriptomic data using Kraken and Mothur', *Frontiers in Microbiology,* 8(2445).

Newbold, C. J., de la Fuente, G., Belanche, A., Ramos-Morales, E. and McEwan, N. R. (2015) 'The Role of Ciliate Protozoa in the Rumen', *Front Microbiol,* 6, 1313.

Nkrumah, J. D., Okine, E. K., Mathison, G. W., Schmid, K., Li, C., Basarab, J. A., Price, M. A., Wang, Z. and Moore, S. S. (2006) 'Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle', *Journal of Animal Science,* 84(1), 145-153.

Nocker, A., Burr, M. and Camper, A. K. (2007) 'Genotypic Microbial Community Profiling: A Critical Technical Review', *Microbial Ecology,* 54(2), 276-289.

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I. V. and Dubchak, I. (2014) 'The genome portal of the Department of Energy Joint Genome Institute: 2014 updates', *Nucleic Acids Research,* 42(Database issue), D26-D31.

NRC (2016) *Nutrient Requirements of Beef Cattle,* 8th Revised Edition ed.*,* Washington, D.C., USA: National Academy Press.

Olfert, E., Cross, B. and McWilliams, A. (1993) *Guide to the Care and Use of Experimental Steers,* Ottawa, ON, Canada: Canadian Council on Animal Care.

Opdahl, L. J., Gonda, M. G. and St-Pierre, B. (2018) 'Identification of Uncultured Bacterial Species from Firmicutes, Bacteroidetes and CANDIDATUS Saccharibacteria as Candidate Cellulose Utilizers from the Rumen of Beef Cows', *Microorganisms,* 6(1).

Orpin, C. G. (1975) 'STUDIES ON RUMEN FLAGELLATE NEOCALLIMASTIX-FRONTALIS', *Journal of General Microbiology,* 91(DEC), 249-262.

Orpin, C. G. and Joblin, K. N. (1997) 'The rumen anaerobic fungi' in Hobson, P. N. and Stewart, C. S., eds., *The Rumen Microbial Ecosystem*, Springer Netherlands, 140-195.

Ozkose, E., Thomas, B. J., Davies, D. R., Griffith, G. W. and Theodorou, M. K. (2001) 'Cyllamyces aberensis gen.nov sp.nov., a new anaerobic gut fungus with branched sporangiophores isolated from cattle', *Canadian Journal of Botany-Revue Canadienne De Botanique,* 79(6), 666-673.

Pajarillo, E. A. B., Chae, J. P., Balolong, M. P., Kim, H. B., Kang, D.-K. and Seo, K.-S. (2015) '- Characterization of the Fecal Microbial Communities of Duroc Pigs Using 16S rRNA Gene Pyrosequencing', - 28(- 4), - 591.

Palackal, N., Lyon, C. S., Zaidi, S., Luginbühl, P., Dupree, P., Goubet, F., Macomber, J. L., Short, J. M., Hazlewood, G. P., Robertson, D. E. and Steer, B. A. (2007) 'A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut', *Applied Microbiology and Biotechnology,* 74(1), 113-124.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. and Tyson, G. W. (2015) 'CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes', *Genome Research,* 25(7), 1043-1055.

Parmar, N. R., Solanki, J. V., Patel, A. B., Shah, T. M., Patel, A. K., Parnerkar, S., Kumar, J. I. and Joshi, C. G. (2014) 'Metagenome of Mehsani buffalo rumen microbiota: an assessment of variation in feed-dependent phylogenetic and functional classification', *J Mol Microbiol Biotechnol,* 24(4), 249-61.

Patil, K. R., Haider, P., Pope, P. B., Turnbaugh, P. J., Morrison, M., Scheffer, T. and McHardy, A. C. (2011) 'Taxonomic metagenome sequence assignment with structured output models', *Nature Methods,* 8, 191.

Patra, A., Park, T., Kim, M. and Yu, Z. (2017) 'Rumen methanogens and mitigation of methane emission by anti-methanogenic compounds and substances', *Journal of Animal Science and Biotechnology,* 8(1), 13.

Paul, S. S., Yu, Z., Bu, D., Xu, J. and Hyde, K. D. (2018) 'A phylogenetic census of global diversity of gut anaerobic fungi and a new taxonomic framework', *Fungal Diversity,* 89(1), 253-266.

Pawlowsky- Glahn, V., Egozcue, J. J. and Tolosana- Delgado, R. (2015) *Modelling and Analysis of Compositional Data*.

Pearson, K. (1897) 'Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs', *Proceedings of the Royal Society of London,* 60(359-367), 489-498.

Petri, R. M., Schwaiger, T., Penner, G. B., Beauchemin, K. A., Forster, R. J., McKinnon, J. J. and McAllister, T. A. (2013) 'Characterization of the Core Rumen Microbiome in Cattle during Transition from Forage to Concentrate as Well as during and after an Acidotic Challenge', *Plos One,* 8(12).

Pickering, N. K., Oddy, V. H., Basarab, J., Cammack, K., Hayes, B., Hegarty, R. S., Lassen, J., McEwan, J. C., Miller, S., Pinares-Patino, C. S. and de Haas, Y. (2015a) 'Animal board invited review: genetic possibilities to reduce enteric methane emissions from ruminants', *animal,* 9(9), 1431-40.

Pickering, N. K., Oddy, V. H., Basarab, J., Cammack, K., Hayes, B., Hegarty, R. S., Lassen, J., McEwan, J. C., Miller, S., Pinares-Patiño, C. S. and de Haas, Y. (2015b) 'Animal board invited review: genetic possibilities to reduce enteric methane emissions from ruminants', *Animal : an international journal of animal bioscience,* 9(9), 1431-1440.

Piro, V. C., Matschkowski, M. and Renard, B. Y. (2017) 'MetaMeta: integrating metagenome analysis tools to improve taxonomic profiling', *Microbiome,* 5(1), 101-101.

Poelaert, C., Nollevaux, G., Boudry, C., Taminiau, B., Nezer, C., Daube, G., Schneider, Y. J., Portetelle, D., Théwis, A. and Bindelle, J. (2017) 'Reducing agent can be omitted in the incubation medium of the batch in vitro fermentation model of the pig intestines', *Animal,* 12(6), 1154-1164.

Poretsky, R., Rodriguez, R. L., Luo, C., Tsementzi, D. and Konstantinidis, K. T. (2014) 'Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics', *PLoS One,* 9(4), e93827.

Poulsen, M., Schwab, C., Borg Jensen, B., Engberg, R. M., Spang, A., Canibe, N., Højberg, O., Milinovich, G., Fragner, L., Schleper, C., Weckwerth, W., Lund, P.,

Schramm, A. and Urich, T. (2013) 'Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen', *Nat Commun,* 4, 1428.

Qi, M., Wang, P., O'Toole, N., Barboza, P. S., Ungerfeld, E., Leigh, M. B., Selinger, L. B., Butler, G., Tsang, A., McAllister, T. A. and Forster, R. J. (2011) 'Snapshot of the Eukaryotic Gene Expression in Muskoxen Rumen-A Metatranscriptomic Approach', *Plos One,* 6(5).

Qi, M., Wang, P., Selinger, L. B., Yanke, L. J., Forster, R. J. and McAllister, T. A. (2011) 'Isolation and characterization of a ferulic acid esterase (Fae1A) from the rumen fungus Anaeromyces mucronatus', *Journal of Applied Microbiology,* 110(5), 1341-1350.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. (2013) 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic Acids Research,* 41(Database issue), D590-D596.

Ramírez, L., Arrizon, J., Sandoval, G., Cardador, A., Bello-Mendoza, R., Lappe, P. and Mateos-Díaz, J. C. (2008) 'A New Microplate Screening Method for the Simultaneous Activity Quantification of Feruloyl Esterases, Tannases, and Chlorogenate Esterases', *Applied Biochemistry and Biotechnology,* 151(2), 711-723.

Rashamuse, K. J., Burton, S. G. and Cowan, D. A. (2007) 'A novel recombinant ethyl ferulate esterase from Burkholderia multivorans', *Journal of Applied Microbiology,* 103(5), 1610-1620.

Ribeiro, G. O., Gruninger, R. J., Badhan, A. and McAllister, T. A. (2016) 'Mining the rumen for fibrolytic feed enzymes', *Animal Frontiers,* 6(2), 20-26.

Ricard, G., McEwan, N. R., Dutilh, B. E., Jouany, J.-P., Macheboeuf, D., Mitsumori, M., McIntosh, F. M., Michalowski, T., Nagamine, T., Nelson, N., Newbold, C. J., Nsabimana, E., Takenaka, A., Thomas, N. A., Ushida, K., Hackstein, J. H. P. and Huynen, M. A. (2006) 'Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment', *BMC Genomics,* 7, 22-22.

Rico, D. E., Preston, S. H., Risser, J. M. and Harvatine, K. J. (2015) 'Rapid changes in key ruminal microbial populations during the induction of and recovery from diet-induced milk fat depression in dairy cows', *British Journal of Nutrition,* 114(3), 358-367.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics,* 26(1), 139-140.

Robinson, M. D. and Oshlack, A. (2010) 'A scaling normalization method for differential expression analysis of RNA-seq data', *Genome Biology,* 11(3), R25.

Robinson, M. D. and Smyth, G. K. (2008) 'Small-sample estimation of negative binomial dispersion, with applications to SAGE data', *Biostatistics,* 9(2), 321-332.

Rodríguez, H., Curiel, J. A., Landete, J. M., de las Rivas, B., de Felipe, F. L., Gómez-Cordovés, C., Mancheño, J. M. and Muñoz, R. (2009) 'Food phenolics and lactic acid bacteria', *International Journal of Food Microbiology,* 132(2), 79-90.

Roehe, R., Dewhurst, R. J., Duthie, C.-A., Rooke, J. A., McKain, N., Ross, D. W., Hyslop, J. J., Waterhouse, A., Freeman, T. C., Watson, M. and Wallace, R. J. (2016) 'Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance', *Plos Genetics,* 12(2).

Rohart, F., Gautier, B., Singh, A. and Lê Cao, K.-A. (2017) 'mixOmics: An R package for 'omics feature selection and multiple data integration', *Plos Computational Biology,* 13(11), 1-19.

Rojas-Downing, M. M., Nejadhashemi, A. P., Harrigan, T. and Woznicki, S. A. (2017) 'Climate change and livestock: Impacts, adaptation, and mitigation', *Climate Risk Management,* 16, 145-163.

Romain, G., Thomas, T. and Margaret, E. K. E. (2009) 'Ecological Conditions Affect Evolutionary Trajectory in a Predator-Prey System', *Evolution,* (3), 641.

Romero-Perez, A., Okine, E. K., McGinn, S. M., Guan, L. L., Oba, M., Duval, S. M., Kindermann, M. and Beauchemin, K. A. (2014) 'The potential of 3-nitrooxypropanol to lower enteric methane emissions from beef cattle', *Journal of Animal Science,* 92(10), 4682-4693.

Rosnow, J. J., Anderson, L. N., Nair, R. N., Baker, E. S. and Wright, A. T. (2017) 'Profiling microbial lignocellulose degradation and utilization by emergent omics technologies', *Critical Reviews in Biotechnology,* 37(5), 626-640.

Russell, J. B. and Rychlik, J. L. (2001) 'Factors that alter rumen microbial ecology', *Science,* 292(5519), 1119-1122.

Sabree, Z. L., Rondon, M. R. and Handelsman, J. (2009) 'Metagenomics' in Schaechter, M., ed. *Encyclopedia of Microbiology (Third Edition)*, Oxford: Academic Press, 622-632.

Saha, B. C. (2000) 'α-l-Arabinofuranosidases: biochemistry, molecular biology and application in biotechnology', *Biotechnology Advances,* 18(5), 403-423.

Saha, B. C. (2003) 'Hemicellulose bioconversion', *Journal of Industrial Microbiology and Biotechnology,* 30(5), 279-291.

Salonen, A., Lahti, L., Salojärvi, J., Holtrop, G., Korpela, K., Duncan, S. H., Date, P., Farquharson, F., Johnstone, A. M., Lobley, G. E., Louis, P., Flint, H. J. and de Vos, W. M. (2014) 'Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men', *ISME Journal: Multidisciplinary Journal of Microbial Ecology,* 8(11), 2218-2230.

San-Juan-Vergara, H., Peña, M., Cadena-Cruz, C., Diaz-Olmos, Y., Hurtado-Gómez, L., Vidal-Orjuela, O., Viasus, D., Cervantes-Acosta, G., Zurek, E., Portnoy, I., Vélez, J. I., Ajami, N. J., Petrosino, J. F., Mogollon, C., Hernández, D., Urruchurtu, I., Di-Ruggiero, P., Sanchez-Sit, S., Guardo-García, E. and Torres, N. (2018) 'A Lachnospiraceae-dominated bacterial signature in the fecal microbiota of HIV-infected individuals from Colombia, South America', *Scientific Reports,* 8(1).

Sanchez, C. (2009) 'Lignocellulosic residues: Biodegradation and bioconversion by fungi', *Biotechnology Advances,* 27(2), 185-194.

Schellenberg, J. J., Verbeke, T. J., Sparling, R., McQueen, P., Krokhin, O. V., Wilkins, J. A., Zhang, X., Alvare, G., Fristensky, B., Levin, D. B., Thallinger, G. G. and Henrissat, B. (2014) 'Enhanced whole genome sequence and annotation of Clostridium stercorarium DSM8532T using RNA-seq transcriptomics and high-throughput proteomics', *BMC Genomics,* 15(1).

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. and Weber, C. F. (2009) 'Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities', *Applied & Environmental Microbiology,* 75(23), 7537-7541.

Schnoes, A. M., Brown, S. D., Dodevski, I. and Babbitt, P. C. (2009) 'Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies', *Plos Computational Biology,* 5(12), e1000605.

Seedorf, H., Kittelmann, S., Henderson, G. and Janssen, P. H. (2014) 'RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments', *PeerJ,* 2.

Sehgal, J. P., Jit, D., Puniya, A. K. and Singh, K. (2008) 'Influence of anaerobic fungal administration on growth, rumen fermentation and nutrient digestion in female buffalo calves', *Journal of Animal and Feed Sciences,* 17(4), 510-518.

Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Hadjithomas, M., Varghese, N. J., Paez-Espino, D., Perry, R., Henderson, G., Creevey, C. J., Terrapon, N., Lapebie, P., Drula, E., Lombard, V., Rubin, E., Kyrpides, N. C., Henrissat, B., Woyke, T., Ivanova, N. N. and Kelly, W. J. (2018) 'Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection', *Nature Biotechnology,* 36(4), 359-367.

Shaani, Y., Zehavi, T., Eyal, S., Miron, J. and Mizrahi, I. (2018) 'Microbiome niche modification drives diurnal rumen community assembly, overpowering individual variability and diet effects', *The ISME Journal*.

Shabat, S. K. B., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., Berg Miller, M. E., White, B. A., Shterzer, N. and Mizrahi, I. (2016) 'Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants', *ISME J*.

Shallom, D. and Shoham, Y. (2003) 'Microbial hemicellulases', *Current Opinion in Microbiology,* 6(3), 219-228.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) 'Cytoscape: A software environment for integrated models of biomolecular interaction networks', *Genome Research,* 13(11), 2498-2504.

Shapiro, E., Biezuner, T. and Linnarsson, S. (2013) 'Single-cell sequencing-based technologies will revolutionize whole-organism science', *Nature Reviews Genetics,* 14, 618.

Silley, P. (1985) 'A note on the pectinolytic enzymes of Lachnospira multiparus', *Journal of Applied Bacteriology,* 58(2), 145-149.

Silverman, J. D., Washburne, A. D., Mukherjee, S. and David, L. A. (2017) 'A phylogenetic transform enhances analysis of compositional microbiota data', *Elife,* 6, e21887.

Sirohi, S. K., Pandey, N., Singh, B. and Puniya, A. K. (2010) 'Rumen methanogens: a review', *Indian Journal of Microbiology,* 50(3), 253-262.

Skene, I. K. and Brooker, J. D. (1995) 'Characterization of Tannin Acylhydrolase Activity in the Ruminal Bacterium Selenomonas Ruminantium', *Anaerobe,* 1(6), 321-327.

Snelling, T. J. and Wallace, R. J. (2017) 'The rumen microbial metaproteome as revealed by SDS-PAGE', *BMC Microbiology,* 17(1), 9.

Sniffen, C. J. and Robinson, P. H. (1987) 'Microbial growth and flow as influenced by dietary manipulations', *Journal of Dairy Science,* 70(2), 425-441.

Sowa, Y. and Berry, R. M. (2008) 'Bacterial flagellar motor', *Quarterly Reviews of Biophysics,* 41(2), 103-132.

Stackebrandt, E. and Goebel, B. M. (1994) 'Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology', *International Journal Of Systematic And Evolutionary Microbiology,* 44(4), 846-849.

Steele, M. A., Penner, G. B., Chaucheyras-Durand, F. and Guan, L. L. (2016) 'Development and physiology of the rumen and the lower gut: Targets for improving gut health', *J Dairy Sci*.

Stephen, A. M. (1983) *The Polysaccharides,* San Diego, CA, USA: Academic.

Stevenson, D. M. and Weimer, P. J. (2007) 'Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR', *Applied Microbiology and Biotechnology,* 75(1), 165-174.

Stewart, C. S., Fonty, G. and Gouet, P. (1988) 'The establishment of rumen microbial communities', *Animal Feed Science and Technology,* 21(2), 69-97.

Stewart, R. D., Warr, A., Watson, M., Auffret, M. D., Dewhurst, R. J., Roehe, R., Wiser, A. H., Press, M. O., Langford, K. W., Liachko, I., Snelling, T. J. and Walker, A. W. (2018) 'Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen', *Nature Communications,* 9(1).

Stuart, D. I., Courcelle, E., M√©toz, F. and Gouet, P. (1999) 'ESPript: analysis of multiple sequence alignments in PostScript', *Bioinformatics,* 15(4), 305-308.

Suen, G., Weimer, P. J., Stevenson, D. M., Aylward, F. O., Boyum, J., Deneke, J., Drinkwater, C., Ivanova, N. N., Mikhailova, N., Chertkov, O., Goodwin, L. A., Currie, C. R., Mead, D. and Brumm, P. J. (2011) 'The Complete Genome Sequence of Fibrobacter succinogenes S85 Reveals a Cellulolytic and Metabolic Specialist',

Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Kultima, J. R., Coelho, L. P. and Arumugam, M. (2013) 'Metagenomic species profiling using universal phylogenetic marker genes', 10(12), 1196-9.

Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Kultima, J. R., Coelho, L. P., Arumugam, M., Tap, J., Nielsen, H. B., Rasmussen, S., Brunak, S., Pedersen, O., Guarner, F., de Vos, W. M., Wang, J., Li, J., Doré, J., Ehrlich, S. D. and Stamatakis, A. (2013) 'Metagenomic species profiling using universal phylogenetic marker genes', *Nature Methods,* 10(12), 1196-1199.

Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., Wu, C. H. and the UniProt, C. (2015) 'UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches', *Bioinformatics,* 31(6), 926-932.

Svartström, O., Alneberg, J., Terrapon, N., Lombard, V., de Bruijn, I., Malmsten, J., Dalin, A.-M., El Muller, E., Shah, P., Wilmes, P., Henrissat, B., Aspeborg, H. and Andersson, A. F. (2017) 'Ninety-nine de novo assembled genomes from the moose (Alces alces) rumen microbiome provide new insights into microbial plant biomass degradation', *The ISME Journal*.

Sylvester, J. T., Karnati, S. K., Yu, Z., Morrison, M. and Firkins, J. L. (2004) 'Development of an assay to quantify rumen ciliate protozoal biomass in cows using real-time PCR', *J Nutr,* 134(12), 3378-84.

Tajima, K., Aminov, R. I., Nagamine, T., Matsui, H., Nakamura, M. and Benno, Y. (2001) 'Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR', *Applied and Environmental Microbiology,* 67(6), 2766-2774.

Tap, J., Furet, J.-P., Bensaada, M., Philippe, C., Roth, H., Rabot, S., Lakhdari, O., Lombard, V., Henrissat, B., Corthier, G., Fontaine, E., Doré, J. and Leclerc, M. (2015) 'Gut microbiota richness promotes its stability upon increased dietary fibre intake in healthy adults', *Environmental Microbiology,* 17(12), 4954-4964.

Tapio, I., Shingfield, K. J., McKain, N., Bonin, A., Fischer, D., Bayat, A. R., Vilkki, J., Taberlet, P., Snelling, T. J. and Wallace, R. J. (2016) 'Oral Samples as Non-Invasive Proxies for Assessing the Composition of the Rumen Microbial Community', *PLoS One,* 11(3), e0151220.

Tapio, I., Snelling, T. J., Strozzi, F. and Wallace, R. J. (2017) 'The ruminal microbiome associated with methane emissions from ruminant livestock', *J Anim Sci Biotechnol,* 8, 7.

Tewoldebrhan, T. A., Appuhamy, J. A. D. R. N., Lee, J. J., Niu, M., Seo, S., Jeong, S. and Kebreab, E. (2017) 'Exogenous β-mannanase improves feed conversion efficiency and reduces somatic cell count in dairy cattle', *Journal of Dairy Science,* 100(1), 244-252.

Theodorou, M. K., Mennim, G., Davies, D. R., Zhu, W.-Y., Trinci, A. P. J. and Brookman, J. L. (2007) 'Anaerobic fungi in the digestive tract of mammalian herbivores and

their potential for exploitation', *Proceedings of the Nutrition Society,* 55(3), 913-926.

Thompson, S. (2015) *The effect of diet type on residual feed intake and the use of infrared thermography as a method to predict efficiency in beef bulls*, unpublished thesis University of Manitoba.

Törrönen, A., Harkki, A. and Rouvinen, J. (1994) 'Three-dimensional structure of endo-1,4-beta-xylanase II from Trichoderma reesei: two conformational states in the active site', *The EMBO Journal,* 13(11), 2493-2501.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R. and Gordon, J. I. (2009) 'A core gut microbiome in obese and lean twins', *Nature,* 457(7228), 480-484.

Tymensen, L. D., Beauchemin, K. A. and McAllister, T. A. (2012) 'Structures of free-living and protozoa-associated methanogen communities in the bovine rumen differ according to comparative analysis of 16S rRNA and mcrA genes', *Microbiology,* 158(7), 1808-1817.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S. and Banfield, J. F. (2004) 'Community structure and metabolism through reconstruction of microbial genomes from the environment', *Nature,* 428, 37.

Van Soest, P. J., Robertson, J. B. and Lewis, B. A. (1991) 'Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition', *Journal of Dairy Science,* 74(10), 3583-3597.

Vanwonterghem, I., Evans, P. N., Parks, D. H., Jensen, P. D., Woodcroft, B. J., Hugenholtz, P. and Tyson, G. W. (2016) 'Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota', *Nature Microbiology,* 1, 16170.

Vardakou, M., Dumon, C., Murray, J. W., Christakopoulos, P., Weiner, D. P., Juge, N., Lewis, R. J., Gilbert, H. J. and Flint, J. E. (2008) 'Understanding the Structural Basis for Substrate and Inhibitor Recognition in Eukaryotic GH11 Xylanases', *Journal Of Molecular Biology,* 375(5), 1293-1305.

Varga, G. A. and Kolver, E. S. (1997) 'Microbial and animal limitations to fiber digestion and utilization', *The Journal Of Nutrition,* 127(5 Suppl), 819S-823S.

Veira, D. M. (1986) 'The role of ciliate protozoa in nutrition of the ruminant', *J Anim Sci,* 63(5), 1547-60.

Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M. G., Apajalahti, J., Tysk, C., Hettich, R. L. and Jansson, J. K. (2008) 'Shotgun metaproteomics of the human distal gut microbiota', *The ISME Journal,* 3, 179.

Větrovský, T. and Baldrian, P. (2013) 'The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses', *Plos One,* 8(2), e57923.

Wallace, R. J., Rooke, J. A., Duthie, C.-A., Hyslop, J. J., Ross, D. W., McKain, N., de Souza, S. M., Snelling, T. J., Waterhouse, A. and Roehe, R. (2014) 'Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle', *Scientific Reports,* 4.

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C.-A., Hyslop, J. J., Ross, D. W., Waterhouse, A., Watson, M. and Roehe, R. (2015) 'The rumen microbial metagenome associated with high methane production in cattle', *BMC Genomics,* 16, 1-14.

Wang, L., Hatem, A., Catalyurek, U. V., Morrison, M. and Yu, Z. (2013) 'Metagenomic Insights into the Carbohydrate-Active Enzymes Carried by the Microorganisms Adhering to Solid Digesta in the Rumen of Cows', *Plos One,* 8(11), 1-11.

Wang, S., Giller, K., Kreuzer, M., Ulbrich, S. E., Braun, U. and Schwarm, A. (2017) 'Contribution of Ruminal Fungi, Archaea, Protozoa, and Bacteria to the Methane Suppression Caused by Oilseed Supplemented Diets', *Frontiers in Microbiology,* 8(1864).

Wang, Y. and McAllister, T. A. (2002) 'Rumen microbes, enzymes and feed digestion - A review', *Asian-Australas J Anim Sci,* 15(11), 1659-1676.

Warton, D. I., Wright, S. T. and Wang, Y. (2012) 'Distance-based multivariate analyses confound location and dispersion effects', *Methods in Ecology and Evolution,* 3(1), 89-101.

Weimer, P. J. (1992) 'Cellulose degradation by ruminal microorganisms', *Critical Reviews in Biotechnology,* 12, 189-223.

Weimer, P. J. (1998) 'Manipulating ruminal fermentation: a microbial ecological perspective', *Journal of Animal Science,* 76(12), 3114-3122.

Weimer, P. J. (2015) 'Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations', *Frontiers in Microbiology,* 6(296).

Weimer, P. J., Lopezguisa, J. M. and French, A. D. (1990) 'EFFECT OF CELLULOSE FINE-STRUCTURE ON KINETICS OF ITS DIGESTION BY MIXED RUMINAL MICROORGANISMS INVITRO', *Applied and Environmental Microbiology,* 56(8), 2421-2429.

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vazquez-Baeza, Y., Birmingham, A., Hyde, E. R. and Knight, R. (2017) 'Normalization and microbial differential abundance strategies depend upon data characteristics', *Microbiome,* 5.

Westerhuis, J. A., van Velzen, E. J. J., Hoefsloot, H. C. J. and Smilde, A. K. (2010) 'Multivariate paired data analysis: multilevel PLSDA versus OPLSDA', *Metabolomics,* 6(1), 119-128.

Wilkens, C., Busk, P. K., Pilgaard, B., Zhang, W.-J., Nielsen, K. L., Nielsen, P. H. and Lange, L. (2017) 'Diversity of microbial carbohydrate-active enzymes in Danish anaerobic digesters fed with wastewater treatment sludge', *Biotechnology for Biofuels,* 10(1), 158.

Wilkinson, T. J., Huws, S. A., Edwards, J. E., Kingston-Smith, A. H., Siu-Ting, K., Hughes, M., Rubino, F., Friedersdorff, M. and Creevey, C. J. (2018) 'CowPI: A Rumen Microbiome Focussed Version of the PICRUSt Functional Inference Software', *Frontiers in Microbiology,* 9(1095).

Williams, A. G. and Coleman, G. S. (1997) 'The rumen protozoa' in Hobson, P. N. and Stewart, C. S., eds., *The Rumen Microbial Ecosystem*, Dordrecht: Springer Netherlands, 73-139.

Wilmes, P. and Bond, P. L. (2004) 'The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms', *Environmental Microbiology,* 6(9), 911-920.

Wimberly, B. T., Ramakrishnan, V. and White, S. W. (1997) 'The structure of ribosomal protein S7 at 1.9 Å resolution reveals a β-hairpin motif that binds double-stranded nucleic acids', *Structure,* 5(9), 1187-1198.

Wolff, S. M., Ellison, M. J., Hao, Y., Cockrum, R. R., Austin, K. J., Baraboo, M., Burch, K., Lee, H. J., Maurer, T., Patil, R., Ravelo, A., Taxis, T. M., Truong, H., Lamberson, W. R., Cammack, K. M. and Conant, G. C. (2017) 'Diet shifts provoke complex and variable changes in the metabolic networks of the ruminal microbiome', *Microbiome,* 5(1), 60.

Wolfger, B., Quinn, C., Torres, G. W., Taylor, M. and Orsel, K. (2016) 'Comparison of feeding behavior between black and red Angus feeder heifers', *Canadian Journal of Animal Science,* 96(3), 404-409.

Wood, D. E. and Salzberg, S. L. (2014) 'Kraken: ultrafast metagenomic sequence classification using exact alignments', *Genome Biology,* 15(3), R46-R46.

Wu, L., Wang, H., Zhang, Z., Lin, R., Zhang, Z. and Lin, W. (2011) 'Comparative Metaproteomic Analysis on Consecutively Rehmannia glutinosa-Monocultured Rhizosphere Soil', *Plos One,* 6(5), e20611.

Wu, S. and Zhang, Y. (2007) 'LOMETS: A local meta-threading-server for protein structure prediction', *Nucleic Acids Research,* 35(10), 3375-3382.

Xu, B., Xu, W., Li, J., Dai, L., Xiong, C., Tang, X., Yang, Y., Mu, Y., Zhou, J., Ding, J., Wu, Q. and Huang, Z. (2015) 'Metagenomic analysis of the Rhinopithecus bieti fecal microbiome reveals a broad diversity of bacterial and glycoside hydrolase profiles related to lignocellulose degradation', *BMC Genomics,* 16(1), 174.

Yang, J., Roy, A. and Zhang, Y. (2013) 'Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment', *Bioinformatics,* 29(20), 2588-2595.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2014) 'The I-TASSER Suite: protein structure and function prediction', *Nature Methods,* 12, 7.

Yang, T., Wei, Z., Friman, V.-P., Xu, Y., Shen, Q., Kowalchuk, G. A. and Jousset, A. (2017) 'Resource availability modulates biodiversity-invasion relationships by altering competitive interactions', *Environmental Microbiology,* 19(8), 2984-2991.

Yi, L., Shi, S., Wang, Y., Huang, W., Xia, Z.-a., Xing, Z., Peng, W. and Wang, Z. (2016) 'Serum Metabolic Profiling Reveals Altered Metabolic Pathways in Patients with Post-traumatic Cognitive Impairments', *Scientific Reports,* 6, 21320.

Yin, Y. B., Mao, X. Z., Yang, J. C., Chen, X., Mao, F. L. and Xu, Y. (2012) 'dbCAN: a web resource for automated carbohydrate-active enzyme annotation', *Nucleic Acids Research,* 40(W1), W445-W451.

Yu, Z. and Morrison, M. (2004) 'Improved extraction of PCR-quality community DNA from digesta and fecal samples', *Biotechniques,* 36(5), 808-812.

Zehavi, T., Probst, M. and Mizrahi, I. (2018) 'Insights Into Culturomics of the Rumen Microbiome', *Frontiers in Microbiology,* 9(1999).

Zhang, Y. (2008) 'I-TASSER server for protein 3D structure prediction', *Bmc Bioinformatics,* 9(1), 40.

Zhang, Y. H. P., Himmel, M. E. and Mielenz, J. R. (2006) 'Outlook for cellulase improvement: Screening and selection strategies', *Biotechnology Advances,* (5), 452.

Zhou, M., Chung, Y. H., Beauchemin, K. A., Holtshausen, L., Oba, M., McAllister, T. A. and Guan, L. L. (2011) 'Relationship between rumen methanogens and methane production in dairy cows fed diets supplemented with a feed enzyme additive', *Journal of Applied Microbiology,* 111(5), 1148-1158.

Zhou, M., Hernandez-Sanabria, E. and Guan, L. L. (2009) 'Assessment of the Microbial Ecology of Ruminal Methanogens in Cattle with Different Feed Efficiencies', *Applied and Environmental Microbiology,* 75(20), 6524-6533.

Zhou, M., Hernandez-Sanabria, E. and Guan, L. L. (2010) 'Characterization of Variation in Rumen Methanogenic Communities under Different Dietary and Host Feed Efficiency Conditions, as Determined by PCR-Denaturing Gradient Gel Electrophoresis Analysis', *Applied and Environmental Microbiology,* 76(12), 3776-3786.

Zhou, M., Hünerberg, M., Beauchemin, K. A., McAllister, T. A., Okine, E. K. and Guan, L. L. (2012) 'Individuality of ruminal methanogen/protozoa populations in beef cattle fed diets containing dried distillers' grain with solubles', *Acta Agriculturae Scandinavica: Section A, Animal Science,* 62(4), 273-288.

Zhou, M., Peng, Y.-J., Chen, Y., Klinger, C. M., Oba, M., Liu, J.-X. and Guan, L. L. (2018) 'Assessment of microbiome changes after rumen transfaunation: implications on improving feed efficiency in beef cattle', *Microbiome,* 6(1), 62-62.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. and Smith, G. M. (2009) 'Mixed Effects Models and Extensions in Ecology with R' in *Mixed Effects Models and Extensions in Ecology with R*.