Rumen Microbiome Associated with Feed Efficiency and Host Genetics in Beef Cattle

by

Fuyong Li

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

© Fuyong Li, 2017

Abstract

In ruminants, evidence is accumulating regarding the associations between rumen microbial taxonomic features and feed efficiency. However, to date, how rumen microbial functional features contribute to the variations in feed efficiency of beef cattle has not been well understood. Moreover, whether the rumen microbiota could be selected and regulated by host genetics still needs to be answered. To fill the knowledge gap, four studies (Chapters 2 - 5) were designed and performed in this thesis.

Chapter 2 aimed to develop a pipeline to identify and quantify the active rumen microbiota using total-RNA-based sequencing (metatranscriptomics), and to compare its outcomes with widely used 16S rRNA/rDNA amplicon sequencing. Taxonomic assessments of metatranscriptomics, 16S rRNA and 16S rDNA amplicon sequencing datasets were performed using a pipeline developed in house. Compared to 16S rRNA/rDNA amplicon sequencing, metatranscriptomics can identify more bacterial and archaeal taxa, and detect more interactions among microbial taxa. These findings validated the feasibility to conduct the taxonomic analysis for the active rumen microbiota using metatranscriptomics.

In Chapter 3, metatranscriptomics was applied to characterize active rumen microbiomes of beef cattle with different feed efficiency (efficient: low residual feed intake [L-RFI]; inefficient: high residual feed intake [H-RFI]). *Lachnospiraceae*, *Lactobacillaceae*, and *Veillonellaceae* were more abundant in H-RFI cattle, and *Methanomassiliicoccale* were more abundant in L-RFI ones ($P < 0.10$). Meanwhile, 32 microbial metabolic pathways and 12 carbohydrate-active enzymes were differentially abundant ($P < 0.05$) between two groups, where most of them were more abundant in H-RFI cattle. These results suggest that rumen microbiomes of inefficient cattle may have higher and more diverse activities than those of efficient cattle.

Chapter 4 was conducted to investigate the associations between the rumen microbiome and feed efficiency (RFI) in various beef cattle breeds. Rumen microbiomes from three breeds (Angus, Charolais, and Kinsella composite hybrid) were profiled using metagenomics and metatranscriptomics. There were distinguishable rumen microbiota and functional potentials among three breeds, but differences of functional activities caused by the breed effect were less apparent. Differential microbial taxonomic and functional features at both DNA and RNA levels were detected between H- and L-RFI cattle; nevertheless, most of them only showed differences between H- and L-RFI animals within a breed, suggesting that there are host genetics and rumen microbiome interactions contributing to the variations in feed efficiency.

In Chapter 5, rumen microbiota from a cohort of 712 beef cattle were assessed using 16S rDNA amplicon sequencing, and results showed that breed, sex, and diet could influence rumen microbiota. Heritability estimation was conducted for microbial taxonomic features using the animal model based on the genomic relationship matrix. It revealed that host genetics affected the alpha- and beta-diversities of rumen microbiota, and the abundance of ~30% of rumen microbial taxa (heritability estimate ≥ 0.15). In addition, 19 SNPs located on 12 bovine chromosomes were found to be associated with 14 rumen microbial taxa. Our study revealed the host genetic effect on the rumen microbial colonization in cattle, highlighting the potential to manipulate the rumen microbiota through genetic selection and breeding.

Overall, findings in this thesis enhanced our understanding on the associations between rumen microbial functional features (at both DNA and RNA levels) and feed efficiency in various beef cattle breeds. Moreover, it provides the evidence that the rumen microbiota is partially shaped by host genetics, which built a theoretical foundation for further manipulating the rumen microbiota using genetic approaches to improve feed efficiency in beef cattle.

Preface

This thesis is an original work by Fuyong Li. The research project, of which this thesis is a part, received research ethics approvals from the Livestock Care Committee of the University of Alberta (AUP00000777 and AUP00000882).

Part of Chapter 1 has been submitted to Journal of Dairy Science as an invited review: Li, F., Neves, A. L. A., Ghoshal, B., & Guan, L. L. (2017), titled "Mining metagenomic and metatranscriptomic data for clues about microbial metabolic functions in ruminants". I was responsible for manuscript writing. Neves, A. L. A. and Ghoshal, B. contributed to manuscript editing. Guan, L. L. contributed to concept formation, manuscript writing and editing.

Chapter 2 of thesis has been published as Li, F., Henderson, G., Sun, X., Cox, F., Janssen, P. H., & Guan L. L. (2016). "Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches". *Front Microbiol, 7,* 987. doi:10.3389/fmicb.2016.00987. I was responsible for experimental design, data generation, data analysis and interpretation, and manuscript writing. Henderson, G. contributed to data interpretation and manuscript writing. Sun, X. contributed to experimental protocol optimization. Cox, F. contributed to DNA isolation and 16S rDNA amplicon sequencing. Janssen, P. H. contributed to data interpretation and manuscript writing. Guan L. L. contributed to experimental design, data analysis and interpretation, and manuscript writing.

Chapter 3 has been published as Li, F., & Guan, L. L. (2017). "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. I was responsible for experimental design, data generation, data analysis and interpretation, and manuscript writing. Guan, L. L. contributed to experimental design, data analysis and interpretation, and manuscript writing.

Dedication

This thesis is dedicated to my beloved family!

Sincerely appreciate your support and encouragement all the time!

谨以此论文献给我挚爱的家人！

感谢你们对我一直以来的支持与鼓励！

Acknowledgements

First of all, I would like to express my earnest appreciation to my supervisor Dr. Leluo Guan! Thank you for giving me your patient guidance since we knew each other in 2010! You consider me as your family and always offer me helpful and useful suggestions on both scientific research and life. You not only taught me how to become a happy and optimistic person, but also taught me how to be a hard-working and flexible researcher! Without your help, I am not able to develop my ability to independently and critically think, which is an essential ability for being a PhD. It is my greatest honor to be your student and to complete my PhD program under your supervision, and all your supports to me will be stored in my heart forever.

At the same time, I am profoundly grateful to Dr. Masahito Oba and Dr. Paul Stothard for being my supervisory committee members. During the last five years, you provide me constructive comments and valuable suggestions on my project, which guaranteed the progress of my work. What's more, Dr. Oba provided me fundamental knowledge on ruminant digestion, metabolism, and nutrition, and Dr. Stothard passed on his helpful skills in bioinformatics to me, which built up a foundation for me to conduct my project. Meanwhile, I also wish to say thanks to my previous committee member Dr. Gemma Henderson. I will never forget your help and suggestions on my study!

Furthermore, I sincerely appreciate Dr. Christopher Creevey's help and suggestions on my project. Many thanks for inviting me to your lab, and shared your bioinformatics pipelines for metagenomics and metatranscriptomics to me. Your contribution is an indispensable part of this project. Moreover, lots of thanks to Dr. Changxi Li for your assistance and help when I did my study to link rumen microorganisms to host genetics. Without your guidance, it could not be successfully moved forward!

Moreover, I would like to deliver my profound appreciation to all previous and current members in Dr. Guan's group: Drs. Weiwu Jin, Mi Zhou, Emma Hernandez-Sanabria, Josue Romao, Guanxiang Liang, Nilusha Malmuthuge, Ou Wang, Duanqin Wu, Diming Wang, Yuwei Ren, Ke Zhao, Huizeng Sun, Bin Yang, Zhi Zhu, as well as Yanhong Chen, Xu Sun, Bibaswan Ghoshal, Rebecca Kong, Eoin O'Hara, Yang Song, André Luis Neves, Anusha Bulumulla, Jitka Hromadkova, and Junhong Liu. It is my great pleasure to work together with you. All of you make our group seem like a big family! In addition, please allow me to say thanks to Drs, Chunyan Zhang, Tianfu Yang, and Feng Zhang due to their helps on the data analysis; and thanks to Thomas Hitch for passing on his knowledge to me, and helping me for my metagenomics and metatranscriptomics data analysis.

I would like to express my gratitude to my friends in Edmonton, Jing Cao, Jinhu Liu, Yu Qian, Lihui Du, Xiaoyan Chen, Yalu Yan, Tingting Ju, Weilan Wang, Xiaoxi Lin, Xiaohong Sun, Guanshi Zhang, Mingliang Jin, Qingbaio Xu. You always share my joys and sorrows during these five years, which makes my stay in Edmonton become an impressive memory in my life.

Last but not least, my dear family, please accept my sincerest and warmest appreciation for your understanding, support and encouragement all the time! Greatest thanks to my wife Yuanyao Chen. Whenever I need you, you are always there for me! My love for you is from the bottom of my heart, and I will spend the rest of my life trying to achieve it. Billions of thanks to my beloved parents Xilong Li and Xiaoling Wang. Thanks a million for your always support! It is understatement to say that I deeply love you, because great love is beyond words. Again, I am truly grateful to all of you!

Table of contents

List of tables

List of figures

List of abbreviations

AA: auxiliary activity

ADG: average daily gain

AMOVA: analysis of molecular variance

ANG: Angus

BTA: *Bos taurus* autosome

CAZyme: carbohydrate-active enzyme

CBM: carbohydrate-binding module

cDNA: complementary DNA

CDs: coding sequences

CE: carbohydrate esterase

CH_4: methane

CHAR: Charolais

cpm: counts per million reads

DMI: dry matter intake

DNA Amplicon-seq: sequencing of PCR amplicons of marker genes (e.g., 16S rDNA)

FCR: feed conversion ratio

FDR: false discovery rate

G matrix: genomic relationship matrix

GH: glycoside hydrolase

GHG: greenhouse gas

GT: glycosyl transferase

GWAS: genome-wide association study

HYB: Kinsella composite hybrid

M-metatranscriptome: mRNA-enriched metatranscriptome

MAF: minor allele frequency

mRNA: messenger RNA

MWT: metabolic weight

NGS: next generation sequencing technology

ORF: open reading frame

OTU: operation taxonomic unit

PCA: principal component analysis

PCoA: principal coordinate analysis

PCR: polymerase chain reaction

PL: pectate lyase

QC: Quality control

qPCR: quantitative PCR

qRT-PCR: quantitative reverse transcription PCR

QTL: quantitative trait locus

RFI: residual feed intake

RIN: RNA integrity number

RNA Amplicon-seq: sequencing of PCR amplicons of marker gene transcripts

RNA-seq: RNA sequencing

rRNA: ribosomal RNA

SNP: single nucleotide polymorphism

T-metatranscriptome: total-RNA-based metatranscriptome

VFA: volatile fatty acid

Chapter 1. Literature review

1.1 Background

In Canada, there are about 10 million beef cattle as of January 2017, including 0.2 million bulls, 3.8 million cows, 1.5 million heifers, 1.2 million steers and 3.3 million calves (Statistics Canada, 2017). Around 1.3 million tonnes of beef are produced every year in Canada, and $33 billion of the Canadian economy is from the beef industry annually (Statistics Canada, 2017). Although beef production indeed brings high quality protein sources (e.g., meat) and economic profits to people, there are still several challenges in the current beef production chain. First of all, livestock production, especially the beef operations, competes for resources with humans, not only for land and water, but also for cereal grains (Eisler *et al.*, 2014; Thornton, 2010). As the global human population is projected to reach 9.15 billion in 2050 (Alexandratos and Bruinsma, 2012), more resources are needed for both human consumption and beef operations in the near future. Secondly, the beef cattle operations may bring negative effects on the environment, such as nitrogen-rich pollutants and methane (CH_4) emissions (Morgavi *et al.*, 2013). For instance, the livestock sector accounts for 18% of global greenhouse gas emissions and 35 - 40% of anthropogenic CH_4 emissions (Steinfeld *et al.*, 2006), which is particularly caused by the enteric fermentation and the manure handling in ruminant operations (Xue *et al.*, 2014). At the same time, CH_4 emissions represent 2 - 12% energy loss from the gross energy intake in cattle (Johnson and Johnson, 1995). These challenges are barriers to achieve more sustainable and efficient beef cattle operations.

One strategy to break through these barriers is to improve feed efficiency of beef cattle, which could increase the food utilization rate, thereby making cattle eat less but produce more or maintain the same production level. For example, an efficient individual could consume 3.77 kg

less diet than an inefficient one per day (Basarab *et al.*, 2003). Improvement of feed efficiency could also decrease negative environmental effects of the beef cattle operations, because efficient cattle produce lower amount of manure and emit ~25% less CH_4 than inefficient ones (Hegarty *et al.*, 2007; Nkrumah *et al.*, 2006). Meanwhile, considering feed related costs represent the biggest outgoing expense (55 - 70%) for the beef industry, the increase in feed efficiency directly ascends profits for beef producers. During a typical finishing period (around 112 - 150 days), it costs $38 - 50 less to feed an efficient beef bull than an inefficient one (Crews, 2005), and a 5% improvement of feed efficiency could bring four times greater economic benefits than a 5% improvement of rate of gain (Government of Alberta, http://www1.agric.gov.ab.ca/).

1.2 Feed efficiency

Feed efficiency is a measure to determine the animal productivity based on the amount of dietary inputs, and for meat-producing animals it could be simply described as the amount of feed consumed divided by the animal weight gain during a certain period (Lamb and Maddock, 2009). As we mentioned above, feed efficiency is one of keys related to the sustainability, environmental effects, and profits of beef cattle production, which is attracting more and more research attention. To measure and determine feed efficiency, several indices have been developed, such as residual feed intake (RFI), feed conversion ratio (FCR), gain: feed ratio (G: F), partial efficiency of growth (PEG), maintenance efficiency. Currently, FCR and RFI are two most frequently used indices for feed efficiency in beef cattle that are summarized as follows:

1.2.1 Feed conversion ratio (FCR)

For meat-producing animals, FCR represents the ratio of feed intake to weight gain during the measurement period, which is one of the most traditional indices for feed efficiency and

calculated as average daily feed intake (dry matter intake; DMI) divided by average daily gain (ADG). According to this definition, low and high FCR indicates high and low feed efficiency, respectively. However, FCR is negatively correlated with ADG (Arthur *et al.*, 2001), suggesting that using FCR as the index to select feed efficient animals (lower FCR) may lead to the selection of individuals with higher mature weight/size and faster growth rate (Santana *et al.*, 2012). This undesired selection increases the maintenance and feed requirements, consequently raises feed costs. Indeed, these undesired outcomes have been observed in the beef industry for last several decades (Government of Alberta, http://www1.agric.gov.ab.ca/), and these larger cattle are not necessary better in terms of production and profitability. In addition, FCR is a gross measurement that does not separate energy inputs for maintenance and growth requirements (Carstens and Tedeschi, 2006). Accordingly, FCR is not an ideal index for feed efficiency in the beef cattle operations.

1.2.2 Residual feed intake (RFI)

RFI, also called net feed efficiency (or net feed intake), is defined as the difference between the actual feed intake and the expected feed intake for growth during a test period (Koch *et al.*, 1963). Individuals with negative/positive RFI values are considered to be efficient/inefficient because they eat less/more than expected. For beef cattle, RFI values are calculated based on DMI, ADG, and MWT (metabolic body weight) over a certain test period (Nkrumah *et al.*, 2006), which take both maintenance and growth requirements into consideration via a linear regression (Basarab *et al.*, 2011). At present, ultrasound back fat has also been incorporated into the RFI calculation to make RFI independent from this trait (Basarab *et al.*, 2011). RFI is a moderately heritable trait with a heritability ranging from 0.26 to 0.43 (Crews, 2005; Nkrumah *et al.*, 2007a). Compared to other indices of feed efficiency (e.g., G: F, FCR, PEG, etc.), RFI is independent of

growth and body weight (Koch *et al.*, 1963). As a result, using RFI is suitable to compare differences among animals at different development stages and/or from different breeds, and selecting improved RFI will not show incidental alterations for growth and body weight, unlike FCR. Therefore, RFI is becoming the preferred measurement of feed efficiency for beef cattle.

However, using RFI as feed efficiency index also has limitations. Firstly, ~50% of steers changed their RFI rankings between growing and finishing stages, even under the same diet (Durunna *et al.*, 2011). This RFI re-ranking indicates that RFI measured during a certain period may not accurately represent feed efficiency through the entire life of beef cattle. Secondly, although relationships between RFI and most of meat quality traits are weak (Ceacero *et al.*, 2016), a previous study still reported negative correlations between RFI and a few meat quality traits (Nkrumah *et al.*, 2007a). Selecting low RFI cattle may cause unfavorable decreases in meat quality traits, such as marbling score (Hoque *et al.*, 2006) and tenderness (McDonagh *et al.*, 2001). Last but not least, to calculate RFI, individual feed intake should be recorded for 9-12 weeks using the automatic feeding measurement system (such as the GrowSafe system) (Moore *et al.*, 2009), which is time-consuming and expensive.

1.2.3 Factors affecting RFI

At least five physiological processes, including feed intake, feed digestion, metabolism, activity, and thermoregulation, contribute to 73% of total RFI variations in ruminants (Herd and Arthur, 2009). The roles of these five physiological processes in feed efficiency are summarized as below. **1) *Feed intake***. Feed intake is related to the maintenance requirement of animals. With the increase of feed intake, the energy for feed digestion and the heat increment level of feeding also ascend. Hence, animals with lower RFI should have lower feed intake, because the energy expended as heat increment and fermentation (HIF) is lower in this case, which has been

confirmed by Nkrumah *et al.* (2007a). **2) *Feed digestion***. Digestibility was negatively correlated with RFI in beef cattle, with the correlation coefficiency r = -0.44 (Herd and Arthur, 2009), suggesting that efficient animals (low RFI) have greater digestibility and substrate availability than inefficient (high RFI) ones. **3) *Body composition and metabolism***. Richardson *et al.* (2001) reported that progenies from high RFI parents had higher whole-body fat but lower whole body protein than offspring of low RFI parents, because the fat deposition rate was faster and some undesirable fat (e.g., subcutaneous fat) were accumulated in high RFI animals. In addition, Herd *et al.* (2004) noted that individuals with low RFI deposited more protein with less degradation and/or turnover than individuals with high RFI. **4) *Activities and feeding patterns***. RFI is positively associated with the animal activity level (Rauw *et al.*, 2006), where the activity included feeding, chewing, ruminating, and locomotion at different speeds (Herd and Arthur, 2009). More active animals have higher maintenance requirement, resulting in higher heat production and lower metabolic efficiency (Herd *et al.*, 2004). Meanwhile, it has been observed that high RFI individuals had longer daily feeding duration, longer head down time, and higher feeding frequency than low RFI ones (Nkrumah *et al.*, 2007b). These suggest that inefficient individuals spend longer time standing and eating than efficient animals, where standing consumes more energy than lying (Richardson and Herd, 2004). **5) *Thermoregulation***. Thermoregulation is highly regulated by the respiration rate, but the direct relationship between the respiration rate and RFI has not been studied in ruminants yet. Instead, the relationship between the heat production and RFI was investigated to indirectly represent the relationship between thermoregulation and RFI (Nkrumah *et al.*, 2006). These authors reported a positive correlation between the heat production and RFI in beef steers, suggesting a higher energy loss via thermoregulation in inefficient animals.

From the above described physiological processes, protein turnover, tissue metabolism and stress (37%), digestibility (10%), HIF (9%), physical activity (9%), body composition (5%), and feeding patterns (2%) have been reported as the main biological factors affecting RFI (Herd and Arthur, 2009). In addition, host genetic elements also contribute to RFI variations. For instance, RFI is a moderately heritable trait with a heritability ranging from 0.26 to 0.43 (Crews, 2005; Nkrumah *et al.*, 2007a). In addition, Barendse *et al.* (2007) detected 161 single nucleotide polymorphisms (SNPs) associated with RFI in cattle, representing 141 genetic regions. These SNPs located in miRNA motifs, promoter sequences, mRNA sequences, and so on, suggesting that several different types of DNA variants contribute to RFI variations. Then Nkrumah *et al.* (2007c) identified 8 significant quantitative trait loci (QTLs; $P < 0.05$) on BTA 1, 5, 7, 8, 12, 16, 17, and 26, as well as several suggestive QTLs ($P < 0.10$) for RFI in beef cattle. After that, Sherman *et al.* (2008) reported that bovine *GHR* (growth hormone receptor), *ghrelin*, and *NPY* (neuropeptide Y) genes had effects on RFI, and Karisa *et al.* (2014) also found other genes (n = 24) associated with RFI in cattle, such as *PARP14* (poly(ADP-ribose) polymerase family member 14) gene and *CAST* (calpastatin) gene. Besides DNA variants, shifts of gene expression have also been linked to RFI. Perkins *et al.* (2014) found expression levels of *NPY*, *RLN3* (relaxin 3), *MC4R* (melanocortin 4 receptor), *GnRH* (gonadotropin releasing hormone), *GnIH* (gonadotropin inhibiting hormone), and *POMC* (pro-opiomelanocortin) genes in arcuate nucleus, and *Leptin* gene in adipose tissue, were both related to RFI. Recently, Kong *et al.* (2016) identified 122 differentially expressed genes from rumen epithelial tissues between high and low RFI steers using RNA-sequencing. All these results confirm that host genetic features both at genomic (DNA) and transcriptomic (RNA) levels contribute to RFI variations.

Moreover, RFI could be influenced by environmental factors, such as seasonality and diet. Mujibi *et al.* (2010) suggested that differences between seasons, in terms of temperature,

humidity, solar radiation, and wind speed, could possibly affect feed intake and thus feed efficiency. The dietary effect on RFI in beef cattle was revealed by Durunna *et al.* (2011), through measuring RFI twice under both growing and finishing diets. They reported that more than 50% individuals switched their RFI ranking when diets were changed. This could be explained by the lack of ability of animals to efficiently adapt to a different feed (Durunna *et al.*, 2011) and/or different diet and host interaction patterns, in which the rumen microbiome plays a role. Indeed, it has been reported that the dietary change affected microbial diversities and activities (Bevans *et al.*, 2005; Petri *et al.*, 2013), which were associated with the RFI re-ranking (Hernandez-Sanabria *et al.*, 2012). As one of the most important links between diet, host, and RFI, better understanding the role of the rumen microbiota is essential.

1.3 Rumen microbiota

Rumen is the largest stomach in ruminants, which is an anaerobic chamber harboring a complex microbial community. The rumen microbiota mainly consists of bacteria, archaea, fungi, ciliated protozoa, and phages (with the concentration up to 10^{11}, 10^{9}, 10^{6}, 10^{6}, 10^{10} per g of digesta or per ml of fluid) (Morgavi *et al.*, 2013), which co-evolves with ruminants. Ruminant hosts provide anaerobic environments to rumen microorganisms for residence and growth, while rumen microorganisms produce fiber-degrading enzymes and equip ruminants with the ability to convert cellulosic plant materials into high protein products for the human consumption. These microorganisms can convert feedstuffs into volatile fatty acids (VFAs, e.g., acetate, propionate, butyrate, etc.), ammonia, microbial protein, vitamins, energy, and so on. VFAs produced by the rumen microbiota could meet 70% of the daily energy requirement for the host (Bergman, 1990). In the meantime, the rumen microbiota is also responsible for producing greenhouse gases (e.g., CH_4, CO_2, etc.). Therefore, rumen is one of microbial ecosystems attracting more and more

research attention nowadays. Exploring compositional and functional characteristics of the rumen microbiota could help us optimize strategies to enhance the rumen microbial fermentation and decrease negative environmental effects, ultimately to achieve more sustainable ruminant operations.

1.3.1 Compositional characteristics of the rumen microbiota

Bacteria are the predominant population in the rumen: 70-80% of them are attached to the feed particles (Minato *et al.*, 1993), 20-30% of them are free-floating in the rumen fluid (McAllister *et al.*, 1994), and approximately 1-2% of them attach to the rumen epithelium (Chen *et al.*, 2011). Through estimating the bacterial population using 16S rDNA-based methods, *Firmicutes* and *Bacteroidetes* were regarded as the most predominant bacterial phyla with both highest abundance and the largest number of species-level operation taxonomic units (OTUs) (Kim *et al.*, 2011). *Lachnospiraceae* and *Prevotellaceae* was the most predominant family belonging to *Firmicutes* and *Bacteroidetes*, respectively (Kittelmann *et al.*, 2013). At the genus level, *Prevotella* was the most abundant group among many ruminant species across a wide geographical range (Henderson *et al.*, 2015).

The majority of rumen archaea could be classified as methanogens that belong to phylum *Euryarchaeota* (Hedderich and Whitman, 2006). Rumen methanogens are found in the rumen fluid, attached to protozoa, within protozoa, and attached to the rumen epithelium (Janssen and Kirs, 2008). Among all rumen methanogens identified using 16S rDNA-based methods, *Methanobrevibacter* was the most abundant genus (61.6%), and there were two major clades *Methanobrevibacter gottschalkii* clade (33.6%) and *Methanobrevibacter ruminantium* clade (27.3%) within this genus; the other two abundant genera were *Methanomicrobium* (14.9%) and *Methanomassiliicoccus* (15.8%) (Janssen and Kirs, 2008; Seedorf *et al.*, 2014).

The cell number of rumen ciliated protozoa is not as high as bacteria, but they account for up to half of the rumen microbial biomass (Russell and Rychlik, 2001). Sequencing PCR amplicons of 18S rDNA from protozoa, revealed that two genera *Entodinium* and *Epidinium* were the most abundant protozoa groups in the rumen (representing 38% and 17% of protozoal sequences, respectively) (Henderson *et al.*, 2015; Kittelmann *et al.*, 2013). In addition to these two most abundant protozoal groups, *Ostracodinium*, *Anoplodinium-Diplodinium*, *Eremoplastron-Diploplastron*, *Eudiplodinium*, *Polyplastron*, *Dasytricha*, and *Isotricha* were all present in more than 70% individuals among different ruminant species (Henderson *et al.*, 2015).

Rumen anaerobic fungi belong to phylum *Neocallimastigomycota*, which either attach to the fibrous plant particles or free swim in the rumen fluid (Gordon and Phillips, 1998). Fungi account for 8 - 20% of the rumen microbial biomass (Tapio *et al.*, 2017). Till now, six rumen fungal genera (i.e., *Anaeromyces*, *Caecomyces*, *Cyllamyces*, *Neocallimastix*, *Orpinomyces*, and *Piromyces*) have been isolated and characterized (Firkins and Yu, 2015; Gruninger *et al.*, 2014b). Recently, through using internal transcribed spacers 1 and 2 (*ITS*1 and *ITS*2) as marker genes, some new fungal groups were detected (Koetschan *et al.*, 2014). Through sequencing PCR amplicons of *ITS*1 genes, Kittelmann *et al.* (2013) reported that *Neocallimastix* (28%), *Piromyces* (20%), and novel clades (including SK1, SK3, and SK4; 16%) were predominant fungal genera in the rumen. However, at the RNA level, *Piromyces*, *Neocallimastix*, and *Orpinomyces* were the most active fungal genera based on the metatranscriptomic data (Qi *et al.*, 2011).

Although there is a high number of virus living in the rumen (10^{10} per g of digesta or per ml of fluid) (Morgavi *et al.*, 2013), the structure of rumen viral community has not been well described till now. According to metagenomic data, *Siphoviridae* (36%), *Myoviridae* (28%), and *Podoviridae* (14%) were identified as the most abundant viral families in the rumen (Berg Miller *et al.*, 2012).

1.3.2 Functions of the rumen microbiota

Bacteria play a major role in the conversion of feed to VFAs and microbial protein in the rumen. Specifically, feed particle attached bacteria mainly contribute to the fiber digestion and are responsible for the majority of endoglucanase activity (Koike *et al.*, 2003), free-floating bacteria in the rumen fluid can initiate the digestion of ingested particles (McAllister *et al.*, 1994), and rumen epithelium attached bacteria involve in oxygen scavenging, recycling of epithelial tissue, and urea hydrolysis (Chen *et al.*, 2011).

Methanogens utilize H_2 and carbon substrates (e.g., CO_2, acetate, methanol, formate, etc.) to produce CH_4 and consequently maintain a low H_2 pressure in the rumen (Hedderich and Whitman, 2006). Efficient H_2 elimination could decrease the inhibitory effect of H_2 on the rumen fermentation and thus increase the fermentation rate. There are three major methanogenesis routes existing in the rumen: CO_2 reduction pathway, C_1 compound conversion pathway, and acetate fermentation pathway (Zhou *et al.*, 2010). Most of rumen methanogens carry out CO_2 reduction pathway: they utilize H_2 and/or formate as electron donors to reduce CO_2 to CH_4. In C_1 compound conversion pathway, methyl groups in methylamines, methanol, and so on, are main substrates for methanogenesis. Some methyl components are firstly oxidized to CO_2 to generate electrons, which are then used to further reduce the methyl components to CH_4. The substrate of acetate fermentation pathway is acetate that could be dissimilated to CH_4 and CO_2 by a few species, but this pathway does not make significant contribution in the rumen (Janssen and Kirs, 2008).

Many ciliated protozoa ingest and store small starch particles, which could adjust the rate of rumen fermentation and avoid rumen acidosis caused by the excessive fermentation (high lactate production and fast drop of pH). Moreover, some protozoa carry out the cellulose digestion that could occupy up to 33% of the rumen fiber digestion capability (Russell and

Rychlik, 2001), and the attachment of protozoa to feed particles could extend their retention time in the rumen. In addition, protozoa are responsible for the bacterial protein turnover through predation of bacteria, and eliminating protozoa could increase the rumen microbial protein supply (Newbold *et al.*, 2015). Last but not least, protozoa are intimately related to methanogenesis due to their high H_2 production ability using hydrogenosomes (Embley *et al.*, 1997; Newbold *et al.*, 2015), and the abundance of protozoa has been correlated to CH_4 emissions (Tapio *et al.*, 2017). However, several defaunation (removing protozoa from the rumen) studies suggest that protozoa are not irreplaceable rumen members for the host survival and nutrition (Newbold *et al.*, 2015).

As initial colonizers of lignocellulosic substrates, rumen fungi could not only produce fibrolytic enzymes to chemically degrade plant cell walls, but also physically penetrate and destroy plant tissues using their rhizoids (Qi *et al.*, 2011). Benefitted from the invasive colonization of fungi to plant tissues, bacteria could more efficiently degrade plant cell walls through the surface erosion, representing a mutualistic relationship between fungi and bacteria in the rumen (France *et al.*, 1990). Through breaking down plant materials, such as cellulose and hemi-cellulose, fungi could generate acetate, propionate, butyrate, as well as other metabolic end products include H_2, CO_2, formate, and so on (Gruninger *et al.*, 2014b; Kittelmann *et al.*, 2012).

Virus could control the amount of microorganisms, select phage-resistant microbes, and promote horizontal gene transfer (HGTs) in an ecosystem (Berg Miller *et al.*, 2012). Because phages are involved in the lysis of bacteria, they probably play an essential role in the dynamics of rumen bacterial population, to maintain the balance of rumen microbial communities (Klieve and Swain, 1993). Considering the lysis of bacteria leads to the destruction of protoplasm, this process may reduce feed conversion efficiency in the rumen. Meanwhile, the lysis of rumen bacteria triggered by phage could be a factor contributing to protein recycling, which also decreases feed utilization efficiency (Swain *et al.*, 1996). What's more, previous studies have

detected transfers of antibiotic resistance genes between bacteria (McCuddin *et al.*, 2006; Toomey *et al.*, 2009), and exchanges of genes coding carbohydrate enzymes between bacteria and eukaryotes (Garcia-Vallve *et al.*, 2000; Ricard *et al.*, 2006). These HGTs could be possibly facilitated by phages (Rohwer and Thurber, 2009).

1.3.3 Factors impacting on the rumen microbiota

The rumen microbiota could be affected by many factors, mainly including diet, feed additives, host species, host age, and physical environmental factors.

1) *Diet*. Diet has been shown to be one of major factors affecting the rumen microbial community in many studies (Ellison *et al.*, 2014; Henderson *et al.*, 2015; Petri *et al.*, 2013; Sun *et al.*, 2010), because it determines available substrates for rumen microorganisms. Specifically, Henderson *et al.* (2015) revealed that diet was the major factor for relative abundance of many rumen bacterial taxa; Ellison *et al.* (2014) also found that rumen microbial communities were distinct between forage-fed and concentrate-fed animals: forage-fed animals had more diverse microbial communities compared to concentrate-fed animals, and rumen methanogens were more prevalent under forage-based diet, which was consistent with results from Petri *et al.* (2013). Through investigating rumen bacterial changes under different concentrate levels, Sun *et al.* (2010) reported that bacterial communities were relatively stable when feeding 0% to 50% concentrates, but high level concentrate diet (70%) indeed caused a decrease in the diversity of rumen bacteria.

2) *Feed additives*. Many feed additives showed a significant influence on the rumen microbiota. Probiotics, such as active dry yeast (ADY), affected the abundance of bacterial taxa (such as SR1, *Fibrobacter succinogenes*, *Prevotella*, etc.) in dairy cows (AlZahal *et al.*, 2017). As a prebiotic, cellooligosaccharides (CE) increased the abundance of fibrolytic bacteria and

methanogenic archaea in dairy calves (Uyeno *et al.*, 2013). Vitamin E did not only lead to higher bacterial and protozoal numbers, but also affected the abundance of some methanogen species according to *in vitro* experiments (Belanche *et al.*, 2016). Fatty acids, such as docosahexaenoic acid (DHA), altered rumen bacterial and archaeal communities, but did not significantly affect rumen fungi or protoza in dairy cows (Torok *et al.*, 2014). In dairy cows, monensin did not affect the archaeal population, but reduced the bacterial diversity, and decreased/increased the abundance of 23/10 bacterial OTUs (Scharen *et al.*, 2017). In addition, medicines, for instance acarbose, changed rumen bacterial community structures, decreased the abundance of *Firmicutes* and *Proteobacteria*, and increased the proportion of *Bacteroidetes*, *Fibrobacteres*, and *Synergistetes* according to in vitro batch culture studies (Yin *et al.*, 2014).

3) *Host species and breed (genetic background)*. Guan *et al.* (2008) reported that rumen bacterial profiles were clustered according to beef cattle breeds. Later on, Hernandez-Sanabria *et al.* (2013) revealed that rumen bacterial communities in the progeny of cattle were affected by the sire breed, and Henderson *et al.* (2015) found that host genetic background affected the abundance of some bacteria groups in the rumen (such as *Fibrobacter*, unclassified *Veillonellaceae*, unclassified *Clostridiales*, and so on).

4) *Host age*. One significant study to link age to the rumen microbiome was conducted in dairy cattle by Jami *et al.* (2013). They found that the abundance of three dominant bacterial phyla *Firmicutes*, *Bacteroidetes* and *Proteobacteria* all shifted along with age. Meanwhile, the diversity of the rumen microbial community increased and the community became more homogeneous as dairy cattle grew.

5) *Physical environmental factors*. Rumen microbial communities are partially structured by several environmental factors, such as ambient temperature (McEwan *et al.*, 2005), day length conditions (McEwan *et al.*, 2005), season (Noel *et al.*, 2017), geographical locations (Ishaq *et al.*,

2015), and so on. In addition, several host physiological features are also related to shifts of the rumen microbiome, including health status (Plaizier *et al.*, 2017), stress (Deng *et al.*, 2017), feed intake (Derakhshani *et al.*, 2016). All these factors mentioned above suggest that the rumen microbiome could be theoretically manipulated through optimal feeding, genetic selection, and improved management for hosts.

1.4 Host genetic effect on the gut microbiota

Due to the effect of host genetic background on the rumen microbiota, it is reasonable to speculate that there are genetic elements (e.g., SNPs, genes, QTLs, etc.) contributing to the variations in the rumen microbial community. Detection of these rumen-microbiota-associated genetic features is the foundation to manipulate rumen microorganisms using genetic selection and breeding. At the same time, identifying these genetic features could help us better understand mechanisms behind interactions between ruminants and their microbiota, which is a fundamental and primary step for optimizing host-microorganism interaction patterns in the future. To date, although no study has been done to identify potential rumen-microbiota-associated genetic elements in ruminants, studies have been conducted in human and mouse to link host genetics and the gut microbiota.

The effect of a single gene on the gut microbiota could be detected using candidate gene methods, in which one gene is added or deleted using an animal model (for example: mouse). Through these approaches, host genes were proved to have significant influence on the diversity and structure of the gut microbiota (Kostic *et al.*, 2013; Spor *et al.*, 2011). For instance, the knockout of *NLRP6* (NOD-like receptor family pyrin domain containing 6) gene increased the abundance of *Bacteroidetes* and TM7 (Elinav *et al.*, 2011), while the knockout of *TLR2* (toll-like receptor 2) gene significantly decreased the abundance of *Firmicutes*, and tended to increase the

abundance of *Proteobacteria*, *Bacteroidetes* and *Actinobacteria* in mouse gut (Kellermayer *et al.*, 2011). In human, it is not feasible to perform gene knock-out or knock-down, but analyzing effects of different genotypes within a gene on the gut microbiota is an alternative way. Khachatryan *et al.* (2008) reported that mutations of *MEFV* (mediterranean fever) gene decreased total bacteria numbers, reduced diversities, shifted the abundance of members belonging to *Bacteroidetes*, *Firmicutes* and *Proteobacteria* in human gut; Frank *et al.* (2011) found that genotypes of *NOD2* (nucleotide-binding oligomerization domain-containing protein 2) and *ATG16L1* (autophagy related 16 like 1) genes were associated with shifts of the gut microbial composition in human.

At present, quantitative genetics, especially genome-wide association study (GWAS), is becoming a powerful strategy to globally scan gut-microbiome-associated genes and QTLs across the whole genome. These approaches have been applied in human and mouse: Benson *et al.* (2010) treated the abundances of gut microorganisms as quantitative and polygenic traits, and identified 18 QTLs significantly or suggestively associated with the mouse gut microbiome. Then a follow-up study reported 42 QTLs for the abundance of 39 microbial taxa in a different mouse strain (Leamy *et al.*, 2014), including 4 QTLs that have been detected previously (Benson *et al.*, 2010). In Goodrich *et al.* (2014) and Goodrich *et al.* (2016), authors assessed heritabilities of gut microbial features in human and found that one-third identified OTUs were heritable with heritability no less than 0.15, providing the evidence that the variations of gut microbiota are affected by the host genetic effect. Hence, it is becoming more certain that there are genetics components associated with the human gut microbiome. Indeed, using data from Human Microbiome Project (HMP), Blekhman *et al.* (2015) reported overall host genetic variations were correlated with the gut microbial structure. They also found that the abundance of *Bifidobacterium* was associated with *LCT* (lactase) and *UBXN4* (UBX domain protein 4) genes,

and *Prevotella* was associated with *PLXND1* (plexin D1), *OR1S1* (olfactory receptor family 1 subfamily S member 1), and *TBPL2* (TATA-box binding protein like 2) genes, as well as the associations between *Alistipes* and *FAM113A* (PC-esterase domain containing 1A) gene, and between *Lachnobacterium* and *EPDR1* (ependymin related 1) gene. Also based on HMP data, Ma *et al.* (2014) revealed that mitochondrial DNA haplogroups were not only associated with the abundance of several gut microbial taxa, but also associated with the general community structure, in which haplogroups refer to groups of similar haplotypes (groups of genes) that share several polymorphisms (Stewart and Chinnery, 2015). At the same time, they reported that there were 5 mtSNPs (mitochondrial single nucleotide polymorphisms) associated with the gut microbial taxonomy profile. Through combining GWAS and the candidate gene approach, Goodrich *et al.* (2016) obtained substantial associations between genes and the gut microbiota in UK population. Specifically, six genes (e.g., *GNA12* [G protein subunit alpha 12], *OR6A2* [olfactory receptor family 6 subfamily A member 2], *CD36* [CD36 molecule], *LCT*, *ALDH1L1* [aldehyde dehydrogenase 1 family member L1], and *RAB3GAP1* [RAB3 GTPase activating protein catalytic subunit 1]) were validated to be associated with the abundance of five gut microbial taxa (SMB53, Cc 115, *Blautia*, *Bifidobacterium*, Unclassified SHA-98) using the candidate gene method. The association between *LCT* gene and *Bifidobacterium* is consistent with the previous study (Blekhman *et al.*, 2015). Then GWAS revealed additional associations: *SLIT3* (slit guidance ligand 3) gene with unclassified *Clostridiaceae*, *R3HDM1* (R3H domain containing 1) gene with *Bifidobacterium*, *UHRF2* (ubiquitin like with PHD and ring finger domains 2) gene with weighted UniFrac distance (beta-diversity index), and two SNPs on chromosome 4 with Bray Curtis dissimilarity (beta-diversity index). Recently, Turpin *et al.* (2016) performed GWAS to evaluate the associations between host genetic variations and the gut microbiota in two human cohorts (discovery cohort [n = 1098] and replication cohort [n = 463]).

In total of 58 SNPs were associated with the abundance of 33 gut microbial taxa in the discovery cohort; 4 of them (located in *CNTN6* [contactin 6], *DMRTB1* [DMRT like family B with proline rich C-terminal 1], *SALL3* [spalt like transcription factor 3], and close to *UBR3* [ubiquitin protein ligase E3 component n-recognin 3] genes) were further confirmed in the replication cohort to be associated with *Faecalibacterium*, *Lachnospira*, *Eubacterium*, and *Rikenellaceae*, respectively.

All these above-mentioned studies verified the feasibility to investigate the association between host genetics and the gut microbiome using GWAS. However, the rumen is different with the post-gastric mammalian digestive tract in terms of physiological characteristics, so that host effects identified within monogastric animals (such as immune system, secreted antimicrobial peptides, and so on) may be less on the rumen microbiome (Henderson *et al.*, 2015). Therefore, it is necessary to perform GWAS on the rumen microbiome, to discover genetics elements contributing to rumen microbial variations.

1.5 Methodologies to study rumen microorganisms

1.5.1 Cultivation-based approaches

For a long time, the rumen microbiota was investigated primarily using cultivation-based approaches, comprising isolation and cultivation of pure strains, which allow us to look at metabolic functions of an isolate and define its ecologic niche in the rumen. Most of our existing knowledge on rumen microbial functions is based on these methods. However, many rumen microorganisms cannot grow by cultivation (Kim *et al.*, 2011), and thus research attention is being moved from cultivation-based methods to culture-independent molecular biology techniques recently. Even so, cultivation-based methods are not out of date. By using these approaches, we could sequence the complete genome of an isolate, predict its functions, and then further verify these predicted functions using physiological studies (Creevey *et al.*, 2014).

Information obtained from cultivation-based methods is the foundation to build reference datasets and to facilitate the culture-independent data analysis. For example, the Hungate 1000 project is on its way to provide a comprehensively rumen-specific database after sequencing around 1000 genomes of rumen bacteria, archaea, fungi, and ciliate protozoa. To date, the Hungate 1000 project has 437 rumen microbial cultures being sequenced, and 311 genomes are already available at the Joint Genome Institute (JGI; http:// genome.jgi.doe.gov).

1.5.2 Marker gene based approaches to study the rumen microbiota

The development of culture-independent molecular techniques, mainly based on marker genes (e.g., 16S rDNA for bacteria and archaea, *mcr*A gene for archaea, 18S rDNA for protozoa, and internal transcribed spacer [*ITS*] gene for fungi, etc.), opened a new window to study the composition of the rumen microbiota. There are different culture-independent approaches that have been applied to study the rumen microbiota: PCR amplification, quantitative real-time PCR, PCR-based fingerprinting methods (PCR-DGGE [denaturing gradient gel electrophoresis], PCR-TTGE [temporal temperature gradient gel electrophoresis], T-RFLP [terminal restriction fragment length polymorphism]), Fluorescence in Situ Hybridisation (FISH), sequencing of clone library, etc. Advantages and limitations of these methods have been comprehensively reviewed and summarized by Zhou *et al.* (2011b). Nowadays, analyzing PCR amplicons of marker genes/transcripts using next generation sequencing technologies (called amplicon sequencing), is a well-accepted approach to generate rumen microbial compositional profiles, which have been applied in several significant studies (Henderson *et al.*, 2015; Kim and Yu, 2014; Kittelmann *et al.*, 2013). However, inherent PCR biases coming from the primer selection (Hong *et al.*, 2009) and amplification cycling conditions (Huber *et al.*, 2009) are unavoidable in amplicon sequencing. In addition, amplicon sequencing has the limitation to discover novel phylotypes,

because PCR primers are usually designed according to known sequences (Ross *et al.*, 2012; Urich *et al.*, 2008).

1.5.3 Metagenomics and metatranscriptomics to study the rumen microbiome

1.5.3.1 Metagenomics and metatranscriptomics

Metagenomics and metatranscriptomics refer to analyzing the assemblage of entire microbial genomes and transcriptomes from environmental samples. Initially, these methodologies started with the cloning of environmental DNA/RNA into vectors (e.g. fosmids, bacterial artificial chromosome vectors, etc.), followed by sequence-based and/or function-driven analysis. Nevertheless, the rapid reduction of costs for NGS accelerated the development of those methodologies, and currently they are usually defined as the direct high-throughput shotgun sequencing of total DNA and RNA in environmental samples (aka shotgun metagenomics and metatranscriptomics), which are adopted in this thesis. The cloning-vector-based metagenomics and metatranscriptomics, as well as related studies will not be included and summarized here.

As mentioned above, metagenomics and metatranscriptomics have become powerful and feasible tools to explore the microbiome of an ecosystem. By using metagenomics and metatranscriptomics, functional potentials (DNA-based) and functional activities (RNA-based) can be estimated, respectively. The advantage of these approaches is that they can better answer the two basic questions: “Which members are there?” and “What are they doing?”.

“Which members are there?” Although amplicon sequencing is rapid and low cost, the taxonomic assessment of the microbiota can be misleading due to inherent PCR biases and limitations to discover novel phylotypes. However, PCR biases can be avoided by skipping the targeted PCR procedure when metagenomics and metatranscriptomics data are used to estimate the taxonomic composition of the microbiota. Recently, it has been reported that metagenomics

and metatranscriptomics can be used for the taxonomic assessment of the microbiota, which is less biased and more quantitative than amplicon sequencing.

"What are they doing?" For a long time, the study of microbial functions highly relied on pure culture-based approaches, which allow us to investigate a specific isolate's metabolic functions and to obtain its single genome/transcriptome. But culture-based methods are not able to characterize functions of uncultured microorganisms, or to elucidate overall functions of all microorganisms within a complex microbiota. Because metagenomics and metatranscriptomics can capture the whole genomic and transcriptomic repertoire for both cultivable and uncultivable microorganisms, they are immensely helpful in functional prediction with high accuracy.

Till now, microbiomes from several ecosystems have been explored using metagenomics and metatranscriptomics, including soil (Tveit *et al.*, 2014; Urich *et al.*, 2008), seawater (Baker *et al.*, 2013; Martinez *et al.*, 2013), animal gastrointestinal tract (Franzosa *et al.*, 2014; Qin *et al.*, 2010), plant rhizosphere (Mendes *et al.*, 2014), and so on. But the number of rumen-related metagenomic and metatranscriptomic studies is still low. Understanding microbial taxonomic and functional characteristics is vital to link the rumen microbiome to host phenotypes, which can help us develop strategies to optimize the rumen microbial fermentation for higher productivity.

1.5.3.2 Application of metagenomics and metatranscriptomics in studying the rumen microbiome

Previous rumen metagenomic and metatranscriptomic studies mainly focused on genes/transcripts encoding enzymes for carbohydrate metabolism, especially enzymes for the degradation of lignocellulosic material that comes from plant cell walls. For instance, Brulc *et al.* (2009) conducted the first shotgun metagenomics-based study to screen carbohydrate-active enzymes (CAZymes), and identified genes encoding glycoside hydrolases (GHs, n = 35),

carbohydrate-binding modules (CBMs, n = 3), carbohydrate esterases (CEs, n = 5), and pectate lyases (PLs, n = 1) from the fiber-adherent rumen microorganisms in beef cattle. Later on, shotgun metagenomics was applied to identify biomass-degrading genes in rumen microorganisms of dairy cattle by Hess *et al.* (2011). In that study, a total of 27,755 carbohydrate-active genes were detected and 90 proteins were expressed; 57% of them were related to the degradation of cellulosic substrates. Meantime, Qi *et al.* (2011) performed a metatranscriptomic analysis to estimate functions of rumen eukaryotes in muskoxen, and reported the expressions of genes encoding GHs, CEs, PLs, and CBMs in eukaryotes of the muskoxen rumen. Later on, a number of follow-up studies have been conducted to detect carbohydrate-degrading enzymes as well as related components in diverse ruminant species using metagenomics and metatranscriptomics (Bensoussan *et al.*, 2017; Comtet-Marre *et al.*, 2017; Dai *et al.*, 2012; Findley *et al.*, 2011; Gruninger *et al.*, 2014a; Jose *et al.*, 2017; Lim *et al.*, 2013; Lopes *et al.*, 2015; Pope *et al.*, 2012).

Recently, metagenomics and metatranscriptomics have been applied to study the rumen virome, whose composition and functions are largely unknown. Berg Miller *et al.* (2012) conducted the first metagenomics-based survey of the rumen virome in dairy cattle and reported that the rumen virome was enriched for phages and transposable elements. According to their results, rumen viruses not only had large numbers, but also displayed high diversity. Later on, Ross *et al.* (2013c) obtained 14 putative viral sequence fragments from dairy cattle rumen metagenomes. They found taxonomically variable but functionally conserved rumen viromes among individual dairy cows. In addition, Yutin *et al.* (2015) explored the diversity of virophages and identified a new family of virophages using metagenomics. The findings from above studies suggest that metagenomics is a powerful tool to discover novel viruses, which opens a new

window to intensively study functions of the rumen virome, and to better understand how viruses interact with other microbial groups (bacteria, archaea, protozoa, and fungi) in the rumen.

In addition to studying the virome, metagenomics and metatranscriptomics have also been used to investigate the rumen bacteria communication pattern via commuquorum-sensing (QS) systems (Ghali *et al.*, 2016), as well as functional potentials of the rumen microorganisms to degrade royal demolition explosives (RDX) (Li *et al.*, 2014). Overall, these highlight the potentials of metagenomics and metatranscriptomics to comprehensively discover novel functions and features of the rumen microbiome.

1.5.4 Pipelines to conduct rumen metagenomics and metatranscriptomics

To date, standard pipelines and general programs for rumen metagenomic and metatranscriptomic analysis have not been well established. Therefore, integrating and comparing outcomes from different studies is still a problem. To solve this problem, in this section, an overview of the major steps of metagenomics and metatranscriptomics is summarized (**Figure 1.1**), including nucleic acid extraction, sequencing platform selection, library construction, quality control of sequencing outputs, compositional analysis, assembly-based functional analysis, assembly-free functional analysis, and comparative analysis.

1.5.4.1 DNA/RNA extraction

To perform metagenomics and metatranscriptomics, the first step is to isolate high quality DNA/RNA from rumen samples. Yu and Morrison (2004) and Henderson *et al.* (2013) compared effects of different DNA extraction methods on rumen samples. Although it is not conclusive which method is universally reliable in all situations, these two studies recommended to use repeated bead beating plus column (RBB+C) or phenol-chloroform and bead beating with

filtration kit (PCQI) methods to isolate DNA from rumen samples. Wang *et al.* (2011) developed an RNA extraction method specifically for rumen samples, which is based on liquid nitrogen grinding of whole ruminal solids and acid guanidinium-phenol-chloroform extraction plus column purification (SRCI). This method was used in the rumen metatranscriptomic study conducted by Qi *et al.* (2011). High quality RNA with RNA integrity number (RIN) higher than 7.0 can be obtained using these rumen-specific RNA extraction protocols, which are recommended for further rumen metatranscriptomic studies. In addition, other RNA isolation methods were also applied in two independent rumen metatranscriptomic studies (Poulsen *et al.*, 2013; Shi *et al.*, 2014), but qualities of isolated RNA were not mentioned and thus are not able to be evaluated.

1.5.4.2 Sequencing platform selection and library construction

To date, there are two commonly used sequencing platforms for metagenomics and metatranscriptomics: Roche/454 platform (shut down in 2013) and Illumina system (https://www.illumina.com), although a few studies have also tried Pacific Biosciences platform (http://www.pacb.com/) (Driscoll *et al.*, 2017; Tsai *et al.*, 2016) and Ion Torrent platform (https://www.thermofisher.com) (Whiteley *et al.*, 2012). Currently, the majority of rumen-related studies are relying on the Illumina system and thus the following sections will focus on this system.

For metatranscirptomic studies, before generating libraries, whether the mRNA enrichment is necessary or not should be considered based on research objectives. Because a large fraction of total rumen microbial RNA constitutes rRNA (> 90%) (Poulsen *et al.*, 2013), the enrichment of mRNA can be conducted to increase the sequencing depth of mRNA in order to capture more active functional transcripts. For example, Shi *et al.* (2014) and AlZahal *et al.*

(2017) successfully enriched mRNA from rumen content/fluid samples using commercial kits. However, Tveit *et al.* (2014) reported that the mRNA enrichment biased the distribution of mRNA, although it increased the proportion of mRNA. Hence, Tveit *et al.* (2014) suggested to omit the mRNA enrichment step and to use total RNA for the library construction, which can provide both taxonomic information (estimated based on rRNA) and functional profiles (derived from mRNA) for the active microbiome. With the reduction in NGS costs, sequencing total RNA with greater depth becomes more feasible.

1.5.4.3 Quality control (QC) of sequencing outputs

After obtaining sequencing reads, assessing the output quality and performing quality control (QC) are crucial steps prior to further analysis, which mainly includes two aspects while studying the rumen microbiome. The first aspect is to trim low quality bases and residual artificial sequences. These noises decrease the accuracy of taxonomic analysis (Bokulich *et al.*, 2013), disturb read mapping, and affect de-novo assembly (Sturm *et al.*, 2016). Several programs have been developed for this step, including Trimmomatic (Bolger *et al.*, 2014), PRINSEQ (Schmieder and Edwards, 2011), FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), NGS QC Toolkit (Patel and Jain, 2012). The main functions of these widely-used programs are summarized in **Table 1.1**. Among these programs, Trimmomatic is a flexible and efficient preprocessing tool that is specifically designed for Illumina outputs and can generate outcomes superior to, or at least competitive with other programs (Bolger *et al.*, 2014), and thus we recommend to use this program for the QC process. In addition to the programs listed in **Table 1.1**, there are other available options that have been briefly compared by Pandey *et al.* (2016), to which the reader is referred.

The second aspect of QC is to remove host sequences, because these sequence contaminations possibly cause mis-assembly of microbial sequences and lead to erroneous conclusions (Schmieder and Edwards, 2011). In human fecal and milk samples, human DNA sequences reached as high as 64% and 77% of total metagenome libraries, respectively, suggesting the contamination of host genetic materials is a serious concern for metagenomic studies (Schmieder and Edwards, 2011; Ward *et al.*, 2013). In rumen meta-omic studies, although host contaminations are not serious (0.43%) (Shabat *et al.*, 2016), checking and discarding them are still recommended. To filter the host (such as *Bos taurus* for cattle and *Ovis aries* for sheep) sequences, reads from rumen metagenomic and metatranscriptomic datasets need to be aligned to the host (*Bos tauras* or *Ovis aries*) genomes and mapped sequences should be discarded. Currently, the MG-RAST (Meyer *et al.*, 2008) based pipeline and the TopHat2 (Kim *et al.*, 2013) based approach have been suggested by Shabat *et al.* (2016) and our research group for this step, respectively. MG-RAST is a web-based data analysis pipeline, while TopHat2 is a program to map/align RNA-Seq reads to reference genomes using a local computer. For this step, the algorithms behind these two methods were both Bowtie-based (Langmead and Salzberg, 2012), and thus the overall outcomes should not be significant different between them. Researchers can make their own choices based on their available computing sources: if they have a local computer with high calculation capacity, they can run the TopHat2 based pipeline; if they don't have a powerful computer, MG-RAST can be a good choice, but the running time after uploading data to MG-RAST is longer than running the analysis locally.

1.5.4.4 Compositional analysis

Amplicon sequencing is the most widely applied approach to estimate the compositional profiles of the rumen microbiota (Henderson *et al.*, 2015; Kittelmann *et al.*, 2013), which has

been mentioned in section **1.5.2**. To achieve the taxonomic analysis of amplicon sequencing, there are mainly three strategies to define Operational Taxonomic Unit (OTU): *de novo*, closed-reference, and open-reference OTU picking (Navas-Molina *et al.*, 2013). Most existing analyses for amplicon sequencing were based on the arbitrary definition of OTUs. Sequences are clustered into OTUs based upon similarity (typically 97% similarity for 16S rDNA sequences), which is called *de novo* OTU picking (Rideout *et al.*, 2014). Although this 97% similarity was considered sufficient to delineate species (Koeppel and Wu, 2013), it is only approximate. Sometimes different species have similar 16S rDNA sequences with high identity (≥ 99%), such as *Bacillus globisporus* and *Bacillus psychrophilus* (Fox *et al.*, 1992), as well as *Clostridium botulinum* and *Clostridium sporogenes* (Rossi-Tamisier *et al.*, 2015). On the contrary, multiple copies of 16S rDNA from a single species may have similarity lower than 97% (Vetrovsky and Baldrian, 2013). Accordingly, the diversity of the microbiota may not be accurately estimated based on these arbitrarily defined OTUs, and analysis at the OTU level may not reflect the "true" and "biological meaningful" microbial composition completely. Therefore, along with the improvement of reference databases, aligning sequences directly to pre-defined reference sequences (called closed-reference OTU picking) (Rideout *et al.*, 2014), is a potential choice to avoid the problem of *de novo* OTU picking. However, it is notable that closed-reference OTU picking cannot identify novel phylotypes, because sequences not matched to references are excluded from the further analysis. Therefore, after closed-reference OTU picking, performing an addition step of *de novo* OTU picking for those sequences failed to match to references (called open-reference OTU picking) can complement this limitation (Rideout *et al.*, 2014), which is recommended for the taxonomic analysis using amplicon sequencing.

In addition to the amplicon sequencing approach mentioned above, taxonomic analysis can also be performed based on metagenomic and metatranscriptomic datasets. From

metagenomic datasets, the taxonomic profiles can be estimated based on metagenomic 16S rDNA reads. This strategy was firstly suggested and tested in marine metagenomes by Logares *et al.* (2014). Later on, Ellison *et al.* (2014) and Taxis *et al.* (2015) applied this approach to evaluate the rumen microbial composition. Due to the low proportion of 16S rDNA among total DNA (0.001% - 0.12%) (Ellison *et al.*, 2014; Logares *et al.*, 2014), the shortcoming of this approach is that high sequencing depth is required to generate enough 16S rDNA reads, which dramatically increases the costs. For metatranscriptomic studies, 16S rRNA extracted from total RNA sequencing can be utilized to generate rumen taxonomic profiles and to represent active microbial taxa. This approach was firstly developed for soil metatranscriptomes by Urich *et al.* (2008), and then was applied to study the active microbiota in soil (Tveit *et al.*, 2014), hydrothermal vents (Lanzen *et al.*, 2011), and animal gut (Poulsen *et al.*, 2013; Schwab *et al.*, 2014). However, it has not been used for rumen samples. To perform these rDNA/rRNA based taxonomic analyses, rDNA/rRNA reads should be identified and extracted from metagenomic and metatranscriptomic datasets, which can be achieved using programs such as SortMeRNA (Kopylova *et al.*, 2012), Bowtie2 (Langmead and Salzberg, 2012), and BLAST (Altschul et al., 1990).

Furthermore, other alternative strategies, such as aligning total metagenome and metatranscriptome reads to genome references or clade-specific marker gene databases, are also used to generate the taxonomic profiles, which have been adopted in the various bioinformatics pipelines including MG-RAST pipeline (Meyer *et al.*, 2008), Phymm/PhymmBL program (Brady and Salzberg, 2009), and MetaPhlAn software (Segata *et al.*, 2012). In addition, Ghoshal *et al.* (unpublished data) recently developed a pipeline to enhance the compositional analysis of rumen metagenome datasets using Kraken (Wood and Salzberg, 2014) and customized databases. However, it is notable that these strategies rely heavily on the integrity of databases, and it is probably not efficient for samples from poorly characterized environments (e.g. rumen).

1.5.4.5 Assembly-based functional analysis

Two assembly strategies. Assembly of short reads is expected to reconstruct genomes and transcriptomes of both cultivable and uncultured organisms from sequencing datasets, which can provide valuable information on protein coding genes and/or operons for the whole microbiome. Despite the impossibility of obtaining complete genome(s) or transcriptome(s) in most cases, assembly is still worth performing in order to obtain longer contigs (long sequence fragments combined from short reads) that will increase the sequence annotation accuracy. For metatranscriptomic data, rRNA sequences should be discarded from the dataset before the assembly, and only the filtered putative mRNA reads should be included into the downstream functional analysis. Two strategies can be employed for assembly: reference-based assembly and *de novo* assembly.

The reference-based assembly strategy only works when sequencing reads are from species with known reference genomes or from species closely related to the known reference genomes. If there are insertions, deletions, or copy number variations, these regions cannot be successfully assembled using reference-based approaches (Thomas *et al.*, 2012). Compared to reference-based assembly, *de novo* assembly reconstructs contigs from sequencing datasets without the aid of genome information. Therefore, considering the complexity of the rumen microbiome and the unavailability of a high number of rumen microbial genomes, *de novo* assembly strategy is primarily recommended in rumen-related metagenomic and metatranscriptomic studies. Currently, most of the existing *de novo* assemblers have been designed based on the *de Bruijn* graph, such as Meta-Velvet (Namiki *et al.*, 2012), SOAPdenovo2 (Luo *et al.*, 2012), Velvet (Zerbino and Birney, 2008), MEGAHIT (Li *et al.*, 2015), and metaSPAdes (Nurk *et al.*, 2017). Till now, the majority of rumen-related metagenomic and metatranscriptomic studies have included the *de novo* assembly step in the

analysis, with only two exceptions (AlZahal *et al.*, 2017; Poulsen *et al.*, 2013). It is noticeable that due to conserved genome regions, organism variability, and gene shifts, *de novo* assembly may bring chimeras and ambiguities into contigs, especially for complex microbial communities (Segata *et al.*, 2013), such as rumen.

Gene prediction and functional annotation. After assembly of metagenomic reads, gene prediction and annotation procedures are similar to the framework of the single microbial genome analysis. Contigs are scanned to identify protein coding sequences (CDs) and/or open reading frames (ORFs), as well as other components. The gene prediction can be achieved using several gene predictors that are particularly designed for metagenomics, including Prodigal (Hyatt *et al.*, 2010), MetaGeneMark (Zhu *et al.*, 2010), FragGeneScan (Rho *et al.*, 2010), and so on. The gene prediction step should be omitted from the metatranscriptomic analysis, and assembled RNA contigs (e.g. transcripts) can be used directly for the annotation. The principle behind the most frequent annotation approach is to assign predicted CDs/ORFs/transcripts against well-characterized databases using homology search, either based on nucleotide or translated protein sequences. There are a number of well-accepted databases for the annotation, such as RefSeq (Pruitt *et al.*, 2012), UniProt (Pundir *et al.*, 2017), KEGG (Kanehisa *et al.*, 2012), eggNOG (Huerta-Cepas *et al.*, 2016), COG/KOG (Tatusov *et al.*, 2003), and SEED (Overbeek *et al.*, 2005). Several programs, including BLAST (Altschul *et al.*, 1990), USEARCH (Edgar, 2010), DIAMOND (Buchfink *et al.*, 2015), are all available to perform homology detection. In addition to the homology-based approach, Hidden Markov Models (HMM)-based method (Eddy, 2011) can also be used to identify protein domains, and are more suitable to annotate sequences with low identity to references. The annotation procedure can also be conducted through web-based annotation pipelines, such as MG-RAST (Meyer *et al.*, 2008), WebMGA (Wu *et al.*, 2011), and Galaxy (https://usegalaxy.org/), which combine several of above-mentioned tools.

Quantification of predicted genes/transcripts. After assembly, sequencing data only have quality characteristics (presence and absence of genes/transcripts) without quantity characteristics (abundance and expressions). To perform the subsequent comparative analysis among treatments and/or samples, original sequencing reads should be mapped to predicted genes/transcripts to estimate the abundance/expressions using tools such as Bowtie2 (Langmead and Salzberg, 2012), BWA (Li and Durbin, 2010), USEARCH (Edgar, 2010), DIAMOND (Buchfink *et al.*, 2015), or BLAST (Altschul *et al.*, 1990). Outputs from these bioinformatics programs are usually SAM (Sequence Alignment/Map), BAM (Binary Alignment/Map), or BLAST-format files. In order to make these outputs adapted for the downstream statistical analysis, they should be further converted to count or abundance files using programs including HUMAnN (Abubucker *et al.*, 2012), MEGAN (Huson *et al.*, 2016), HTSeq (Anders *et al.*, 2015), and SAMtools (Li *et al.*, 2009a).

1.5.4.6 Assembly-free functional analysis

The assembly-free approach is to align original short reads (post QC) to existing functional databases or annotated reference genomes, which is considered as an alternative way to estimate functional profiles of metagenomes and metatranscriptomes. For metatranscriptomic data, rRNA reads should be removed and only putative mRNA sequences should be kept for the analysis. This approach has been used in many metagenomic and metatranscriptomic studies (Dubin *et al.*, 2016; Franzosa *et al.*, 2014; Sanders *et al.*, 2015), but only two of them are rumen-related (AlZahal *et al.*, 2017; Poulsen *et al.*, 2013). The accuracy of the assembly-free approach highly depends on the reference databases. For well-characterized environments, such as human gut with 77% of the taxa cultured and genomes identified (Lagier *et al.*, 2016), this approach is reliable and its outcomes are convincing (Dubin *et al.*, 2016; Franzosa *et al.*, 2014). However, for

poorly-described environments without complete references, such as rumen (with > 55% "unclassified" and "unnamed" microbial phylotypes) (Henderson *et al.*, 2015), only a low proportion of mRNA reads could be mapped to the KEGG database (~8%; Li *et al.*, unpublished data.) and the remaining unmapped reads have to be ignored for the downstream analysis. This can explain why only two rumen-related meta-omics studies adopted this approach (AlZahal *et al.*, 2017; Poulsen *et al.*, 2013).

MG-RAST (Meyer *et al.*, 2008), HUMAnN (Abubucker *et al.*, 2012), and MEGAN (Huson *et al.*, 2016) are well-accepted programs to conduct the assembly-free approach. The feasibilities of these programs have been tested, and they have been successfully applied to study microbiomes from many environments, such as soil (Urich *et al.*, 2008), animal gut (Franzosa *et al.*, 2014; Sanders *et al.*, 2015), and plant rhizosphere (Mendes *et al.*, 2014). Therefore, we suggest that these programs can be applied to study the rumen microbiome if the assembly-free approach is to be used. With the Hungate 1000 project on its way, it will provide a comprehensively rumen-specific database after sequencing around 1000 genomes of rumen bacteria, archaea, fungi, and ciliate protozoa, which will make the assembly-free approach more feasible for characterization of rumen metagenomes and metatranscriptomes in the near future. To date, the Hungate 1000 project has 437 rumen microbial cultures being sequenced, and 311 genomes are already available at the Joint Genome Institute (JGI; http:// genome.jgi.doe.gov).

1.5.4.7 Comparative analysis

In the majority of metagenomic and metatranscriptomic studies, the ultimate goal is to identify differences among treatments and/or samples at taxonomic and functional levels. An important aspect before the comparative analysis is to apply the appropriate normalization to all related counts (e.g., genes, transcripts, functional pathways, etc.) obtained from previous steps.

This is because different samples have different sequencing depths and genes/transcripts are not of the same length. To normalize different sequencing depths, the counts per million reads (cpm) is commonly used by dividing the count of each gene/transcript to the total number of mapped reads. The limitation of this approach is that the length variations of genes/transcripts are not accounted for, and thus it is suitable to compare the same functional features between samples but not suitable to compare different functional features within the same sample. If characterizing the functional profile and comparing different genes/transcripts within the same sample is one of the research objectives, it is necessary to take the length of each gene/transcript into consideration and normalize the gene/transcript counts to reads/fragments per kilo base per million mapped reads (RPKM/FPKM). In addition, since most of sequencing data are still not normally distributed after normalization, further data transformation is necessary to make the data more appropriate for frequently-used normal-distribution-based statistical methods, such as ANOVA/t-test based on arcsine square root transformation (Franzosa *et al.*, 2014). Otherwise, approaches specifically-designed for sequencing data should be adopted to perform the comparative analysis, including LEfSe (Segata *et al.*, 2011), DESeq2 (Love *et al.*, 2014), edgeR (Robinson and Smyth, 2007), and so on.

1.6 Linkages between the rumen microbiome and host phenotypes

As an essential part of the digestive track in ruminants, the rumen microbiome (at both taxonomic and functional levels) has been linked to many host phenotypes, such as milk composition (Jami *et al.*, 2014), rumen acidosis (Khafipour *et al.*, 2009), feed efficiency (Hernandez-Sanabria *et al.*, 2010; Shabat *et al.*, 2016), CH_4 yields (Shi *et al.*, 2014; Wallace *et al.*, 2015), and so on.

1.6.1 Feed efficiency

Most of known linkages between feed efficiency and the rumen microbiome were revealed at the taxonomic level. Guan *et al.* (2008) observed low-RFI cattle had more similar rumen bacterial profiles that separated from high-RFI individuals, which is the first attempt to link the rumen microbiome to feed efficiency within our best knowledge. After that, many follow-up studies were conducted to further investigate these linkages and to identify particular rumen microbial taxa related to feed efficiency. The abundance of bacteria taxa at different taxonomic levels, such as phylum-level taxa (*Firmicutes* and *Lentisphaerae*), family-level taxa (*Lachnospiraceae*, *Paraprevotellaceae*, *Prevotellaceae,* RF39, S24-7, and *Veillonellaceae*), genus-level taxa (*Anaerovibrio*, *Butyrivibrio*, *Clostridium*, *Coprococcus*, *Eubacterium*, *Fibrobacter*, *Lactobacillus*, *Prevotella*, *Pseudobutyrivibrio*, *Ruminococcus*, *Succiniclasticum*, and *Succinivibrio*), and so on, are all reported to be associated with feed efficiency in beef and dairy cattle (Carberry *et al.*, 2012; Hernandez-Sanabria *et al.*, 2010; Hernandez-Sanabria *et al.*, 2013; Jewell *et al.*, 2015; Myer *et al.*, 2015a). The abundance of rumen archaeal taxa, including *Methanobrevibacter smithii*, *Methanosphaera stadtmanae*, and *Methanobrevibacter* sp. strain AbM4, as well as the overall archaeal community structure, are also associated with feed efficiency in beef cattle (Carberry *et al.*, 2014; Zhou *et al.*, 2009; Zhou *et al.*, 2010).

However, at the functional level, the associations between the rumen microbiome and feed efficiency have not been well estimated and more research attention is being paid on this aspect. Ross *et al.* (2012) verified the variations of rumen metagenomes between different dairy cows were larger than the variations caused by repeated samplings within the same cow, firstly indicating the feasibility to link the rumen metagenome to host phenotypes. Applying metagenomics to rumen samples from dairy cows, Shabat *et al.* (2016) revealed that more efficient individuals possess lower diverse rumen microbiomes at both species and gene levels,

and the prediction of feed efficiency using genes of rumen microbiomes can reach up to 91% accuracy. In the meantime, they found that genes enriched in the efficient group mostly belonged to *Megasphaera elsdenii*, while the similar trend was not observed in the inefficient group. Therefore, they suggested that microbiomes of inefficient cattle might utilize more diverse substrates and produce more diverse products, which either have negative effect on the host or cannot be efficiently absorbed by the host. These findings suggest that rumen microbiomes of inefficient cattle have more diverse functional potentials (at the DNA level) than those of efficient cattle, which contribute to the host feed efficiency variations.

1.6.2 CH_4 emissions

As major CH_4 producers, the overall abundance of total rumen methanogens only had weak or even no correlation with CH_4 emission levels, while the composition and the structure of the archaeal community had more significant effects on CH_4 yield (Tapio *et al.*, 2017). For example, the relative abundance of *Methanobrevibacter gottschalkii*, *Methanogenic archaeon*, *Methanosphaera* spp., *Methanomassiliicoccus*, and so on contribute to the variations of CH_4 emissions in cattle and sheep (Kittelmann *et al.*, 2014; Poulsen *et al.*, 2013; Shi *et al.*, 2014; Zhou *et al.*, 2011a). In addition to archaeal communities, rumen bacterial profiles are also associated with CH_4 emissions (Kittelmann *et al.*, 2014): there are three different ruminotypes comprising various bacterial taxa, in which higher abundance of H_2-producing bacteria correlated with higher CH_4 yield. The abundance of *Proteobacteria*, especially family *Succinivibrionaceae*, showed a negative correlation with CH_4 emissions (Wallace *et al.*, 2015). Furthermore, the abundance of rumen protozoa was reported to be positively correlated with CH_4 yield using a meta-analysis (Guyader *et al.*, 2014), and this negative correlation pattern was confirmed through the elimination of rumen ciliate protozoa that reduced 11% CH_4 production (Newbold *et al.*, 2015).

The first attempt to link the rumen metagenome to CH_4 emissions was conducted in dairy cows by Ross *et al.* (2013b), who successfully predicted CH_4 yields using metagenomic profiles with an accuracy of prediction 0.466. Soon afterwards, Shi *et al.* (2014) reported that it was the expressions of methanogenesis-related genes instead of the abundance of these genes contributing to the CH_4 yield variations in sheep, when sheep rumen microbiomes were assessed using metagenomics and metatranscriptomics. Specifically, the CH_4 metabolism pathway (ko00680), three genes encoding subunits of methyl coenzyme M reductase, and several genes encoding enzymes for the CO_2/H_2 pathway expressed higher in high CH_4 emitters. Therefore, they suggested that reducing expression levels of these genes could be a strategy for future CH_4 mitigation. Contrary to that study, Wallace *et al.* (2015) revealed genes directly or indirectly involved in methanogenesis were on average 2.7-fold more abundant in high CH_4 yield beef cattle based on rumen metagenomics, suggesting that functional characteristics of the rumen microbiome at the DNA level should also be considered as the target for future CH_4 mitigation. Through combining metagenomics, metatranscriptomics, and amplicon sequencing, Kamke *et al.* (2016) reported that *Sharpea* spp. and *Megasphaera* spp. were more abundant, and genes/transcripts for sugar transport, lactate utilisation and production, propionate metabolism, and butyrate metabolism were enriched in low CH_4 yield sheep. These suggest that more hexoses were firstly fermented to lactate and then converted to butyrate by *Sharpea* spp. and *Megasphaera* spp. through a two-step process in the rumen of those low CH_4 yield animals. In addition, through comparing microbiomes between yaks/Tibetan sheep from high-altitude (low CH_4 producers) and cattle/ordinary sheep from low altitude (high CH_4 producers) using metagenomics, Zhang *et al.* (2016) found that ruminants with low CH_4 emissions had more abundant energy- and carbohydrate-metabolic-related functional categories, while high CH_4 producers had enriched functions related to methanogenesis (such as the CO_2/H_2 pathway and the

methylotrophic pathway). Overall, these suggest that metagenomics and metatrnascriptomics can reveal the potentials (DNA) and activities (RNA) of rumen functions and/or enzymes involved in methanogensis, providing future targets for methane mitigation in ruminants.

1.6.3 Microbial responses to diet and feed additives

As described in section **1.3.2**, diet and feed additive are two main factors to shape the rumen microbial community structures. As one of the main factors to shape the rumen microbiome (Henderson *et al.*, 2015), diet contributes to rumen microbial variations at both taxonomic and functional levels, which has been confirmed using metagenomics (Campanaro *et al.*, 2017; Ellison *et al.*, 2014). Metagenomics was also applied to study the rumen metabolic disorder (e.g. frothy bloat) caused by microbial functional shifts after beef cattle received the high crude protein (CP) diet (Pitta *et al.*, 2016).

Metagenomics and metatranscriptomics were utilized to look into effects of feed additives on the rumen microbiome. For example, Ross *et al.* (2013a) investigated the effect of two methane-mitigating feed additives (grapemarc and a combination of lipids and tannin) on the rumen microbiome using metagenomics in dairy cattle. These two additives actually altered microbiomes in similar ways, both of which affected four second level KEGG categories (e.g. amino acid metabolism, carbohydrate metabolism, translation, and biosynthesis of other secondary metabolites). Meanwhile, adding rapeseed oil (RSO) supplementation to the diet of dairy cows reduced CH_4 emissions, and the mechanism of this effect was investigated using rumen metatranscriptomics (Poulsen *et al.*, 2013). After adding RSO to diet, it was observed that down-regulated expressions in several microbial genes of enzymes involved in methylamine-based methanogenesis, including dimethylamine permease (DMA), DMA methyltransferase corrinoid protein, trimethylamine:corrinoid methyltransferase, which were all related to the

decrease in *Methanomassiliicoccales*. Denman *et al.* (2015) studied the CH_4 inhibition effect of bromochloromethane (BCM, an anti-methanogenic compound) on goats via metagenomics. Higher abundance of genes involved in the succinate pathway were detected in rumen microbiomes in response to BCM, and thus they speculated that this pathway could primarily consume H_2 when methanogenesis was inhibited. The above findings indicate that metagenomics and metatranscriptomics are power tools to detect the microbial shifts in response to dietary intervention strategies.

1.7 Knowledge gaps

Although rumen microorganisms have been explored in the last decade using culture-independent molecular techniques, the current understanding of its phylogenetic diversities is primarily from studies using marker genes based approaches as well as a few studies using metagenomics. Because marker genes based approaches have inherent biases as we mentioned in section **1.5.2** and DNA-based metagenomics could not distinguish active, inactive, and dead cells, metatranscriptomics is becoming a potentially powerful approach to identify active microorganisms, which has been applied in soil (Tveit *et al.*, 2014; Urich *et al.*, 2008), hydrothermal vents (Lanzen *et al.*, 2011), animal gut (Poulsen *et al.*, 2013; Schwab *et al.*, 2014), etc. However, metatranscriptomics has not been used to assess the compositional profiles of the rumen microbiota and there is no well-defined pipeline to achieve the taxonomic analysis based on rumen metatranscriptomic data. Moreover, our understanding of the associations between functional features of the rumen microbiome and feed efficiency is merely based on a recent metagenomic study in dairy cattle (Shabat *et al.*, 2016). So far, no study has been conducted to link rumen microbial functional features (at both DNA and RNA levels) to feed efficiency in beef cattle, and whether these associations are universal among different beef cattle breeds is

unknown. Furthermore, it has been known that several genes and QTLs contribute to the variations of the gut microbiota in humans and mice (section **1.4**), and host genetic background could impact on the rumen microbiota (Guan *et al.*, 2008; Henderson *et al.*, 2015; Hernandez-Sanabria *et al.*, 2013). However, mechanisms and genetic components behind the host genetic effect have not been investigated in ruminants.

1.8 Hypotheses and objectives

I hypothesized that total-RNA-based metatranscriptomics could be applied to characterize both compositional and functional characteristics of the rumen microbiome, and metatranscriptomes differ among cattle with varied feed efficiency and different breeds. Moreover, I also hypothesized that rumen microorganisms can be influenced by host genetics, and bovine genetic elements (e.g., SNPs, genes, QTLs, etc.) can contribute to the variations of the rumen microbiota.

The long-term objectives of this project were to identify associations between the rumen microbiome and feed efficiency, and to detect bovine genetic components for the rumen microbiome that are responsible for higher feed efficiency. The specific objectives were: **1)** to develop a pipeline to achieve the taxonomic analysis using metatranscriptomic data (Chapter 2); **2)** to assess compositional and functional profiles of the active rumen microbiome using metatranscriptomics, and to link these active microbial features with feed efficiency in beef cattle (Chapter 3); **3)** to investigate associations between rumen metagenome/metatranscriptome and feed efficiency (RFI) in various beef cattle breeds (Chapter 4); **4)** to identify bovine genetic components (e.g., SNPs, genes, QTLs, etc.) for the rumen microbiota using GWAS (Chapter 5).

1.9 References

Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL *et al* (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* **8:** e1002358.

Alexandratos N, Bruinsma J (2012). World agriculture towards 2030/2050: the 2012 revision. ESA Working paper Rome, FAO.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

AlZahal O, Li F, Guan LL, Walker ND, McBride BW (2017). Factors influencing ruminal bacterial community diversity and composition and microbial fibrolytic enzyme abundance in lactating dairy cows with a focus on the role of active dry yeast. *J Dairy Sci*.

Anders S, Pyl PT, Huber W (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:** 166-169.

Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, Parnell PF (2001). Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. *J Anim Sci* **79:** 2805-2811.

Baker BJ, Sheik CS, Taylor CA, Jain S, Bhasi A, Cavalcoli JD *et al* (2013). Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling. *Isme j* **7:** 1962-1973.

Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W, Thomas MB (2007). A validated whole-genome association study of efficient food conversion in cattle. *Genetics* **176:** 1893-1905.

Basarab JA, Price MA, Aalhus JL, Okine EK, Snelling WM, Lyle KL (2003). Residual feed intake and body composition in young growing cattle. *Can J Anim Sci* **83:** 189-204.

Basarab JA, Colazo MG, Ambrose DJ, Novak S, McCartney D, Baron VS (2011). Residual feed intake adjusted for backfat thickness and feeding frequency is independent of fertility in beef heifers. *Can J Anim Sci* **91:** 573-584.

Belanche A, Kingston-Smith AH, Newbold CJ (2016). An Integrated Multi-Omics Approach Reveals the Effects of Supplementing Grass or Grass Hay with Vitamin E on the Rumen Microbiome and Its Function. *Front Microbiol* **7:** 905.

Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J *et al* (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107:** 18933-18938.

Bensoussan L, Morais S, Dassa B, Friedman N, Henrissat B, Lombard V *et al* (2017). Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome. *Environ Microbiol* **19:** 185-197.

Berg Miller ME, Yeoman CJ, Chia N, Tringe SG, Angly FE, Edwards RA *et al* (2012). Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environ Microbiol* **14:** 207-227.

Bergman EN (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev* **70:** 567-590.

Bevans DW, Beauchemin KA, Schwartzkopf-Genswein KS, McKinnon JJ, McAllister TA (2005). Effect of rapid or gradual grain adaptation on subacute acidosis and feed intake by feedlot cattle. *J Anim Sci* **83:** 1116-1132.

Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT *et al* (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biol* **16:** 191.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R *et al* (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* **10:** 57-59.

Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114-2120.

Brady A, Salzberg SL (2009). Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods* **6:** 673-676.

Brulc JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC, Dinsdale EA *et al* (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci U S A* **106:** 1948-1953.

Buchfink B, Xie C, Huson DH (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12:** 59-60.

Campanaro S, Treu L, Cattani M, Kougias PG, Vendramin V, Schiavon S *et al* (2017). In vitro fermentation of key dietary compounds with rumen fluid: A genome-centric perspective. *Sci Total Environ* **584-585:** 683-691.

Carberry CA, Kenny DA, Han S, McCabe MS, Waters SM (2012). Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. *Appl Environ Microbiol* **78:** 4949-4958.

Carberry CA, Waters SM, Kenny DA, Creevey CJ (2014). Rumen methanogenic genotypes differ in abundance according to host residual feed intake phenotype and diet type. *Appl Environ Microbiol* **80:** 586-594.

Carstens GE, Tedeschi LO (2006). Defining feed efficiency in beef cattle. *Proceedings of Beef Improvement Federation 38th Annual Research Symposium and Annual Meeting*; 2006; Choctaw, Mississippi.

Ceacero TM, Mercadante ME, Cyrillo JN, Canesin RC, Bonilha SF, de Albuquerque LG (2016). Phenotypic and Genetic Correlations of Feed Efficiency Traits with Growth and Carcass Traits in Nellore Cattle Selected for Postweaning Weight. *PLoS One* **11:** e0161366.

Chen Y, Penner GB, Li M, Oba M, Guan LL (2011). Changes in bacterial diversity associated with epithelial tissue in the beef cow rumen during the transition to a high-grain diet. *Appl Environ Microbiol* **77:** 5770-5781.

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. *Front Microbiol* **8:** 67.

Creevey CJ, Kelly WJ, Henderson G, Leahy SC (2014). Determining the culturability of the rumen bacterial microbiome. *Microb Biotechnol* **7:** 467-479.

Crews DH, Jr. (2005). Genetics of efficient feed utilization and national cattle evaluation: a review. *Genet Mol Res* **4:** 152-165.

Dai X, Zhu Y, Luo Y, Song L, Liu D, Liu L *et al* (2012). Metagenomic insights into the fibrolytic microbiome in yak rumen. *PLoS One* **7:** e40430.

Deng L, He C, Zhou Y, Xu L, Xiong H (2017). Ground transport stress affects bacteria in the rumen of beef cattle: A real-time PCR analysis. *Anim Sci J* **88:** 790-797.

Denman SE, Martinez Fernandez G, Shinkai T, Mitsumori M, McSweeney CS (2015). Metagenomic analysis of the rumen microbial community following inhibition of methane formation by a halogenated methane analog. *Front Microbiol* **6:** 1087.

Derakhshani H, Tun HM, Cardoso FC, Plaizier JC, Khafipour E, Loor JJ (2016). Linking Peripartal Dynamics of Ruminal Microbiota to Dietary Changes and Production Parameters. *Front Microbiol* **7:** 2143.

Driscoll CB, Otten TG, Brown NM, Dreher TW (2017). Towards long-read metagenomics: complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture. *Stand Genomic Sci* **12:** 9.

Dubin K, Callahan MK, Ren B, Khanin R, Viale A, Ling L *et al* (2016). Intestinal microbiome analyses identify melanoma patients at risk for checkpoint-blockade-induced colitis. *Nat Commun* **7:** 10391.

Durunna ON, Mujibi FDN, Goonewardene L, Okine EK, Basarab JA, Wang Z *et al* (2011). Feed efficiency differences and reranking in beef steers fed grower and finisher diets. *Journal of Animal Science* **89:** 158-167.

Eddy SR (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol* **7:** e1002195.

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-2461.

Eisler MC, Lee MR, Tarlton JF, Martin GB, Beddington J, Dungait JA *et al* (2014). Agriculture: Steps to sustainable livestock. *Nature* **507:** 32-34.

Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ *et al* (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145:** 745-757.

Ellison MJ, Conant GC, Cockrum RR, Austin KJ, Truong H, Becchi M *et al* (2014). Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. *DNA Res* **21:** 115-125.

Embley TM, Horner DA, Hirt RP (1997). Anaerobic eukaryote evolution: hydrogenosomes as biochemically modified mitochondria? *Trends Ecol Evol* **12:** 437-441.

Findley SD, Mormile MR, Sommer-Hurley A, Zhang XC, Tipton P, Arnett K *et al* (2011). Activity-based metagenomic screening and biochemical characterization of bovine ruminal protozoan glycoside hydrolases. *Appl Environ Microbiol* **77:** 8106-8113.

Firkins JL, 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:** 1450-1470.

Fox GE, Wisotzkey JD, Jurtshuk P, Jr. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42:** 166-170.

France J, Theodorou MK, Davies D (1990). Use of zoospore concentrations and life cycle parameters in determining the population of anaerobic fungi in the rumen ecosystem. *J Theor Biol* **147:** 413-422.

Frank DN, Robertson CE, Hamm CM, Kpadeh Z, Zhang T, Chen H *et al* (2011). Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* **17:** 179-184.

Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM *et al* (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* **111:** E2329-2338.

Garcia-Vallve S, Romeu A, Palau J (2000). Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Mol Biol Evol* **17:** 352-361.

Ghali I, Shinkai T, Mitsumori M (2016). Mining of luxS genes from rumen microbial consortia by metagenomic and metatranscriptomic approaches. *Anim Sci J* **87:** 666-673.

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R *et al* (2014). Human genetics shape the gut microbiome. *Cell* **159:** 789-799.

Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C *et al* (2016). Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe* **19:** 731-743.

Gordon GL, Phillips MW (1998). The role of anaerobic gut fungi in ruminants. *Nutr Res Rev* **11:** 133-168.

Gruninger RJ, Gong X, Forster RJ, McAllister TA (2014a). Biochemical and kinetic characterization of the multifunctional beta-glucosidase/beta-xylosidase/alpha-arabinosidase, Bgxa1. *Appl Microbiol Biotechnol* **98:** 3003-3012.

Gruninger RJ, Puniya AK, Callaghan TM, Edwards JE, Youssef N, Dagar SS *et al* (2014b). Anaerobic fungi (phylum Neocallimastigomycota): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiol Ecol* **90:** 1-17.

Guan LL, Nkrumah JD, Basarab JA, Moore SS (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol Lett* **288:** 85-91.

Guyader J, Eugene M, Noziere P, Morgavi DP, Doreau M, Martin C (2014). Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. *Animal* **8:** 1816-1825.

Hedderich R, Whitman WB (2006). Physiology and biochemistry of the methane-producing Archaea. *The prokaryotes*. Springer. pp 1050-1079.

Hegarty RS, Goopy JP, Herd RM, McCorkell B (2007). Cattle selected for lower residual feed intake have reduced daily methane production. *J Anim Sci* **85:** 1479-1486.

Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, Noel SJ *et al* (2013). Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One* **8:** e74787.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH (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.

Herd RM, Oddy VH, Richardson EC (2004). Biological basis for variation in residual feed intake in beef cattle. 1. Review of potential mechanisms. *Australian Journal of Experimental Agriculture* **44:** 423-430.

Herd RM, Arthur PF (2009). Physiological basis for residual feed intake. *J Anim Sci* **87:** E64-71.

Hernandez-Sanabria E, Guan LL, Goonewardene LA, Li M, Mujibi DF, Stothard P *et al* (2010). Correlation of particular bacterial PCR-denaturing gradient gel electrophoresis patterns with bovine ruminal fermentation parameters and feed efficiency traits. *Appl Environ Microbiol* **76:** 6338-6350.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Durunna ON, Moore SS, Guan LL (2012). Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. *Appl Environ Microbiol* **78:** 1203-1214.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Zhou M, Moore SS, Guan LL (2013). Influence of sire breed on the interplay among rumen microbial populations inhabiting the rumen liquid of the progeny in beef cattle. *PLoS One* **8:** e58461.

Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G *et al* (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331:** 463-467.

Hong S, Bunge J, Leslin C, Jeon S, Epstein SS (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *Isme j* **3:** 1365-1373.

Hoque MA, Arthur PF, Hiramoto K, Oikawa T (2006). Genetic relationship between different measures of feed efficiency and its component traits in Japanese Black (Wagyu) bulls. *Livestock Science* **99:** 111-118.

Huber JA, Morrison HG, Huse SM, Neal PR, Sogin ML, Mark Welch DB (2009). Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ Microbiol* **11:** 1292-1302.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC *et al* (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44:** D286-293.

Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S *et al* (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol* **12:** e1004957.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11:** 119.

Ishaq SL, Sundset MA, Crouse J, Wright AG (2015). High-throughput DNA sequencing of the moose rumen from different geographical locations reveals a core ruminal methanogenic archaeal diversity and a differential ciliate protozoal diversity. *Microb Genom* **1:** e000034.

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

Jami E, White BA, Mizrahi I (2014). Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *PLoS One* **9:** e85423.

Janssen PH, Kirs M (2008). Structure of the archaeal community of the rumen. *Appl Environ Microbiol* **74:** 3619-3625.

Jewell KA, McCormick CA, Odt CL, Weimer PJ, Suen G (2015). Ruminal Bacterial Community Composition in Dairy Cows Is Dynamic over the Course of Two Lactations and Correlates with Feed Efficiency. *Appl Environ Microbiol* **81:** 4697-4710.

Johnson KA, Johnson DE (1995). Methane emissions from cattle. *J Anim Sci* **73:** 2483-2492.

Jose VL, More RP, Appoothy T, Arun AS (2017). In depth analysis of rumen microbial and carbohydrate-active enzymes profile in Indian crossbred cattle. *Syst Appl Microbiol* **40:** 160-170.

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:** 56.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40:** D109-114.

Karisa B, Moore S, Plastow G (2014). Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle. *Anim Sci J* **85:** 374-387.

Kellermayer R, Dowd SE, Harris RA, Balasa A, Schaible TD, Wolcott RD *et al* (2011). Colonic mucosal DNA methylation, immune response, and microbiome patterns in Toll-like receptor 2-knockout mice. *Faseb j* **25:** 1449-1460.

Khachatryan ZA, Ktsoyan ZA, Manukyan GP, Kelly D, Ghazaryan KA, Aminov RI (2008). Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS One* **3:** e3064.

Khafipour E, Li S, Plaizier JC, Krause DO (2009). Rumen microbiome composition determined using two nutritional models of subacute ruminal acidosis. *Appl Environ Microbiol* **75:** 7115-7124.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14:** R36.

Kim M, Morrison M, Yu Z (2011). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol Ecol* **76:** 49-63.

Kim M, Yu Z (2014). Variations in 16S rRNA-based microbiome profiling between pyrosequencing runs and between pyrosequencing facilities. *J Microbiol* **52:** 355-365.

Kittelmann S, Naylor GE, Koolaard JP, Janssen PH (2012). A proposed taxonomy of anaerobic fungi (class neocallimastigomycetes) suitable for large-scale sequence-based community structure analysis. *PLoS One* **7:** e36866.

Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI *et al* (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS One* **8:** e47879.

Kittelmann S, Pinares-Patino CS, Seedorf H, Kirk MR, Ganesh S, McEwan JC *et al* (2014). Two different bacterial community types are linked with the low-methane emission trait in sheep. *PLoS One* **9:** e103171.

Klieve AV, Swain RA (1993). Estimation of ruminal bacteriophage numbers by pulsed-field gel electrophoresis and laser densitometry. *Appl Environ Microbiol* **59:** 2299-2303.

Koch RM, Swiger LA, Chambers D, Gregory KE (1963). Efficiency of Feed Use in Beef Cattle. *J Anim Sci* **22:** 486-494.

Koeppel AF, Wu M (2013). Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. *Nucleic Acids Res* **41:** 5175-5188.

Koetschan C, Kittelmann S, Lu J, Al-Halbouni D, Jarvis GN, Muller T *et al* (2014). Internal transcribed spacer 1 secondary structure analysis reveals a common core throughout the anaerobic fungi (Neocallimastigomycota). *PLoS One* **9:** e91928.

Koike S, Pan J, Kobayashi Y, Tanaka K (2003). Kinetics of In Sacco Fiber-Attachment of Representative Ruminal Cellulolytic Bacteria Monitored by Competitive PCR. *Journal of Dairy Science* **86:** 1429-1435.

Kong RS, Liang G, Chen Y, Stothard P, Guan le L (2016). Transcriptome profiling of the rumen epithelium of beef cattle differing in residual feed intake. *BMC Genomics* **17:** 592.

Kopylova E, Noe L, Touzet H (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28:** 3211-3217.

Kostic AD, Howitt MR, Garrett WS (2013). Exploring host-microbiota interactions in animal models and humans. *Genes Dev* **27:** 701-718.

Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P *et al* (2016). Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol* **1:** 16203.

Lamb GC, Maddock T (2009). Feed efficiency in cows. *Cash, Cows, and Calves* **1:** 35-42.

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357-359.

Lanzen A, Jorgensen SL, Bengtsson MM, Jonassen I, Ovreas L, Urich T (2011). Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiol Ecol* **77:** 577-589.

Leamy LJ, Kelly SA, Nietfeldt J, Legge RM, Ma F, Hua K *et al* (2014). Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biol* **15:** 552.

Li D, Liu CM, Luo R, Sadakane K, Lam TW (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31:** 1674-1676.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078-2079.

Li H, Durbin R (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26:** 589-595.

Li RW, Giarrizzo JG, Wu S, Li W, Duringer JM, Craig AM (2014). Metagenomic insights into the RDX-degrading potential of the ovine rumen microbiome. *PLoS One* **9:** e110505.

Lim S, Seo J, Choi H, Yoon D, Nam J, Kim H *et al* (2013). Metagenome Analysis of Protein Domain Collocation within Cellulase Genes of Goat Rumen Microbes. *Asian-Australas J Anim Sci* **26:** 1144-1151.

Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento H *et al* (2014). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* **16:** 2659-2671.

Lopes LD, de Souza Lima AO, Taketani RG, Darias P, da Silva LR, Romagnoli EM *et al* (2015). Exploring the sheep rumen microbiome for carbohydrate-active enzymes. *Antonie Van Leeuwenhoek* **108:** 15-30.

Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J *et al* (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1:** 18.

Ma J, Coarfa C, Qin X, Bonnen PE, Milosavljevic A, Versalovic J *et al* (2014). mtDNA haplogroup and single nucleotide polymorphisms structure human microbiome communities. *BMC Genomics* **15:** 257.

Martinez A, Ventouras LA, Wilson ST, Karl DM, Delong EF (2013). Metatranscriptomic and functional metagenomic analysis of methylphosphonate utilization by marine bacteria. *Front Microbiol* **4:** 340.

McAllister TA, Bae HD, Jones GA, Cheng KJ (1994). Microbial attachment and feed digestion in the rumen. *Journal of Animal Science* **72:** 3004-3018.

McCuddin ZP, Carlson SA, Rasmussen MA, Franklin SK (2006). Klebsiella to Salmonella gene transfer within rumen protozoa: implications for antibiotic resistance and rumen defaunation. *Vet Microbiol* **114:** 275-284.

McDonagh MB, Herd RM, Richardson EC, Oddy VH, Archer JA, Arthur PF (2001). Meat quality and the calpain system of feedlot steers following a single generation of divergent selection for residual feed intake. *Australian Journal of Experimental Agriculture* **41:** 1013-1021.

McEwan NR, Abecia L, Regensbogenova M, Adam CL, Findlay PA, Newbold CJ (2005). Rumen microbial population dynamics in response to photoperiod. *Lett Appl Microbiol* **41:** 97-101.

Mendes LW, Kuramae EE, Navarrete AA, van Veen JA, Tsai SM (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. *Isme j* **8:** 1577-1587.

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

Minato H, Mitsumori M, Cheng KJ (1993). Attachment of microorganisms to solid substrates in the rumen. *Genetics, Biochemistry and Ecology of Lignocellulose Degradation (Shimada, K, Ohmiya, K, Kobayashi, Y, Hoshino, S, Sakka, K and Karita, S, Eds)***:** 139-145.

Moore SS, Mujibi FD, Sherman EL (2009). Molecular basis for residual feed intake in beef cattle. *J Anim Sci* **87:** E41-47.

Morgavi DP, Kelly WJ, Janssen PH, Attwood GT (2013). Rumen microbial (meta)genomics and its application to ruminant production. *Animal* **7 Suppl 1:** 184-201.

Mujibi FD, Moore SS, Nkrumah DJ, Wang Z, Basarab JA (2010). Season of testing and its effect on feed intake and efficiency in growing beef cattle. *J Anim Sci* **88:** 3789-3799.

Myer PR, Smith TP, Wells JE, Kuehn LA, Freetly HC (2015). Rumen microbiome from steers differing in feed efficiency. *PLoS One* **10:** e0129174.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* **40:** e155.

Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z *et al* (2013). Advancing Our Understanding of the Human Microbiome Using QIIME. *Methods in Enzymology*. pp 371-444.

Newbold CJ, de la Fuente G, Belanche A, Ramos-Morales E, McEwan NR (2015). The Role of Ciliate Protozoa in the Rumen. *Front Microbiol* **6:** 1313.

Nkrumah JD, Okine EK, Mathison GW, Schmid K, Li C, Basarab JA *et al* (2006). Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle. *J Anim Sci* **84:** 145-153.

Nkrumah JD, Basarab JA, Wang Z, Li C, Price MA, Okine EK *et al* (2007a). Genetic and phenotypic relationships of feed intake and measures of efficiency with growth and carcass merit of beef cattle. *J Anim Sci* **85:** 2711-2720.

Nkrumah JD, Crews DH, Jr., Basarab JA, Price MA, Okine EK, Wang Z *et al* (2007b). Genetic and phenotypic relationships of feeding behavior and temperament with performance, feed efficiency, ultrasound, and carcass merit of beef cattle. *J Anim Sci* **85:** 2382-2390.

Nkrumah JD, Sherman EL, Li C, Marques E, Crews DH, Jr., Bartusiak R *et al* (2007c). Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. *J Anim Sci* **85:** 3170-3181.

Noel SJ, Attwood GT, Rakonjac J, Moon CD, Waghorn GC, Janssen PH (2017). Seasonal changes in the digesta-adherent rumen bacterial communities of dairy cattle grazing pasture. *PLoS One* **12:** e0173819.

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27:** 824-834.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M *et al* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33:** 5691-5702.

Pandey RV, Pabinger S, Kriegner A, Weinhausel A (2016). ClinQC: a tool for quality control and cleaning of Sanger and NGS data in clinical research. *BMC Bioinformatics* **17:** 56.

Patel RK, Jain M (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **7:** e30619.

Perkins SD, Key CN, Garrett CF, Foradori CD, Bratcher CL, Kriese-Anderson LA *et al* (2014). Residual feed intake studies in Angus-sired cattle reveal a potential role for hypothalamic gene expression in regulating feed efficiency. *J Anim Sci* **92:** 549-560.

Petri RM, Schwaiger T, Penner GB, Beauchemin KA, Forster RJ, McKinnon JJ *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:** e83424.

Pitta DW, Pinchak WE, Indugu N, Vecchiarelli B, Sinha R, Fulford JD (2016). Metagenomic Analysis of the Rumen Microbiome of Steers with Wheat-Induced Frothy Bloat. *Front Microbiol* **7:** 689.

Plaizier JC, Li S, Danscher AM, Derakshani H, Andersen PH, Khafipour E (2017). Changes in Microbiota in Rumen Digesta and Feces Due to a Grain-Based Subacute Ruminal Acidosis (SARA) Challenge. *Microb Ecol*.

Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC *et al* (2012). Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PLoS One* **7:** e38571.

Poulsen M, Schwab C, Jensen BB, Engberg RM, Spang A, Canibe N *et al* (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat Commun* **4:** 1428.

Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* **40:** D130-135.

Pundir S, Martin MJ, O'Donovan C (2017). UniProt Protein Knowledgebase. *Methods Mol Biol* **1558:** 41-55.

Qi M, Wang P, O'Toole N, Barboza PS, Ungerfeld E, Leigh MB *et al* (2011). Snapshot of the eukaryotic gene expression in muskoxen rumen-A metatranscriptomic approach. *PLoS ONE* **6**.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C *et al* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464:** 59-65.

Rauw WM, Soler J, Tibau J, Reixach J, Gomez Raya L (2006). The relationship between residual feed intake and feed intake behavior in group-housed Duroc barrows. *J Anim Sci* **84:** 956-962.

Rho M, Tang H, Ye Y (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* **38:** e191.

Ricard G, McEwan NR, Dutilh BE, Jouany JP, 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.

Richardson EC, Herd RM, Oddy VH, Thompson JM, Archer JA, Arthur PF (2001). Body composition and implications for heat production of Angus steer progeny of parents selected for and against residual feed intake. *Australian Journal of Experimental Agriculture* **41:** 1065-1072.

Richardson EC, Herd RM (2004). Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. *Australian Journal of Experimental Agriculture* **44:** 431-440.

Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM *et al* (2014). Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* **2:** e545.

Robinson MD, Smyth GK (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23:** 2881-2887.

Rohwer F, Thurber RV (2009). Viruses manipulate the marine environment. *Nature* **459:** 207-212.

Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM *et al* (2012). High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* **13:** 53.

Ross EM, Moate PJ, Marett L, Cocks BG, Hayes BJ (2013a). Investigating the effect of two methane-mitigating diets on the rumen microbiome using massively parallel sequencing. *J Dairy Sci* **96:** 6030-6046.

Ross EM, Moate PJ, Marett LC, Cocks BG, Hayes BJ (2013b). Metagenomic predictions: from microbiome to complex health and environmental phenotypes in humans and cattle. *PLoS One* **8:** e73056.

Ross EM, Petrovski S, Moate PJ, Hayes BJ (2013c). Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. *BMC Microbiol* **13:** 242.

Rossi-Tamisier M, Benamar S, Raoult D, Fournier PE (2015). Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. *Int J Syst Evol Microbiol* **65:** 1929-1934.

Russell JB, Rychlik JL (2001). Factors that alter rumen microbial ecology. *Science* **292:** 1119-1122.

Sanders JG, Beichman AC, Roman J, Scott JJ, Emerson D, McCarthy JJ *et al* (2015). Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. *Nat Commun* **6:** 8285.

Santana MHA, Rossi P, Almeida R, Cucco DC (2012). Feed efficiency and its correlations with carcass traits measured by ultrasound in Nellore bulls. *Livestock Science* **145:** 252-257.

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. *J Dairy Sci* **100:** 2765-2783.

Schmieder R, Edwards R (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27:** 863-864.

Schwab C, Berry D, Rauch I, Rennisch I, Ramesmayer J, Hainzl E *et al* (2014). Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *Isme j* **8:** 1101-1114.

Seedorf H, Kittelmann S, Henderson G, Janssen PH (2014). RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. *PeerJ* **2:** e494.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS *et al* (2011). Metagenomic biomarker discovery and explanation. *Genome Biol* **12:** R60.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9:** 811-814.

Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C (2013). Computational meta'omics for microbial community studies. *Mol Syst Biol* **9:** 666.

Shabat SK, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Berg Miller ME *et al* (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *Isme j*.

Sherman EL, Nkrumah JD, Murdoch BM, Li C, Wang Z, Fu A *et al* (2008). Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. *J Anim Sci* **86:** 1-16.

Shi W, Moon CD, Leahy SC, Kang D, Froula J, Kittelmann S *et al* (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res* **24:** 1517-1525.

Spor A, Koren O, Ley R (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* **9:** 279-290.

Steinfeld H, Gerber P, Wassenaar TD, Castel V, de Haan C (2006). *Livestock's long shadow: environmental issues and options*. Food & Agriculture Org.

Stewart JB, Chinnery PF (2015). The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet* **16:** 530-542.

Sturm M, Schroeder C, Bauer P (2016). SeqPurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics* **17:** 208.

Sun YZ, Mao SY, Zhu WY (2010). Rumen chemical and bacterial changes during stepwise adaptation to a high-concentrate diet in goats. *Animal* **4:** 210-217.

Swain RA, Nolan JV, Klieve AV (1996). Natural variability and diurnal fluctuations within the bacteriophage population of the rumen. *Appl Environ Microbiol* **62:** 994-997.

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

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV *et al* (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:** 41.

Taxis TM, Wolff S, Gregg SJ, Minton NO, Zhang C, Dai J *et al* (2015). The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Nucleic Acids Res* **43:** 9600-9612.

Thomas T, Gilbert J, Meyer F (2012). Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp* **2:** 3.

Thornton PK (2010). Livestock production: recent trends, future prospects. *Philos Trans R Soc Lond B Biol Sci* **365:** 2853-2867.

Toomey N, Monaghan A, Fanning S, Bolton D (2009). Transfer of antibiotic resistance marker genes between lactic acid bacteria in model rumen and plant environments. *Appl Environ Microbiol* **75:** 3146-3152.

Torok VA, Percy NJ, Moate PJ, Ophel-Keller K (2014). Influence of dietary docosahexaenoic acid supplementation on the overall rumen microbiota of dairy cows and linkages with production parameters. *Can J Microbiol* **60:** 267-275.

Tsai YC, Conlan S, Deming C, Segre JA, Kong HH, Korlach J *et al* (2016). Resolving the Complexity of Human Skin Metagenomes Using Single-Molecule Sequencing. *MBio* **7:** e01948-01915.

Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI *et al* (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat Genet* **48:** 1413-1417.

Tveit AT, Urich T, Svenning MM (2014). Metatranscriptomic analysis of arctic peat soil microbiota. *Appl Environ Microbiol* **80:** 5761-5772.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3:** e2527.

Uyeno Y, Kawashima K, Hasunuma T, Wakimoto W, Noda M, Nagashima S *et al* (2013). Effects of cellooligosaccharide or a combination of cellooligosaccharide and live Clostridium butyricum culture on performance and intestinal ecology in Holstein calves fed milk or milk replacer. *Livestock Science* **153:** 88-93.

Vetrovsky T, Baldrian P (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* **8:** e57923.

Wallace RJ, Rooke JA, McKain N, Duthie CA, Hyslop JJ, Ross DW *et al* (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* **16:** 839.

Wang P, Qi M, Barboza P, Leigh MB, Ungerfeld E, Selinger LB *et al* (2011). Isolation of high-quality total RNA from rumen anaerobic bacteria and fungi, and subsequent detection of glycoside hydrolases. *Can J Microbiol* **57:** 590-598.

Ward TL, Hosid S, Ioshikhes I, Altosaar I (2013). Human milk metagenome: a functional capacity analysis. *BMC Microbiol* **13:** 116.

Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B *et al* (2012). Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods* **91:** 80-88.

Wood DE, Salzberg SL (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15:** R46.

Wu S, Zhu Z, Fu L, Niu B, Li W (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* **12:** 444.

Xue B, Wang LZ, Yan T (2014). Methane emission inventories for enteric fermentation and manure management of yak, buffalo and dairy and beef cattle in China from 1988 to 2009. *Agriculture, Ecosystems and Environment* **195:** 202-210.

Yin YY, Liu YJ, Zhu WY, Mao SY (2014). Effects of Acarbose Addition on Ruminal Bacterial Microbiota, Lipopolysaccharide Levels and Fermentation Characteristics In vitro. *Asian-Australas J Anim Sci* **27:** 1726-1735.

Yu Z, Morrison M (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36:** 808-812.

Yutin N, Kapitonov VV, Koonin EV (2015). A new family of hybrid virophages from an animal gut metagenome. *Biol Direct* **10:** 19.

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18:** 821-829.

Zhang Z, Xu D, Wang L, Hao J, Wang J, Zhou X *et al* (2016). Convergent Evolution of Rumen Microbiomes in High-Altitude Mammals. *Curr Biol* **26:** 1873-1879.

Zhou M, Hernandez-Sanabria E, Le LG (2009). Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Appl Environ Microbiol* **75:** 6524-6533.

Zhou M, Hernandez-Sanabria E, Guan LL (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. *Appl Environ Microbiol* **76:** 3776-3786.

Zhou M, Chung YH, Beauchemin KA, Holtshausen L, Oba M, McAllister TA *et al* (2011a). Relationship between rumen methanogens and methane production in dairy cows fed diets supplemented with a feed enzyme additive. *J Appl Microbiol* **111:** 1148-1158.

Zhou M, McAllister TA, Guan LL (2011b). Molecular identification of rumen methanogens: Technologies, advances and prospects. *Animal Feed Science and Technology* **166-167:** 76-86.

Zhu W, Lomsadze A, Borodovsky M (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38:** e132.

1.10 Tables

Table 1.1 Widely-used programs for quality control (QC) of sequencing reads

Program	Input	Major QC functions
Trimmomatic (Bolger *et al.*, 2014)	Single-end & Paired-end (Illumina)	1. Cut primer, adapter and other artificial sequences 2. Drop low quality reads 3. Trim low quality bases 4. Crop reads to certain length
PRINSEQ (Schmieder and Edwards, 2011)	Single-end & Paired-end (Illumina, 454/Roche, Ion Torrent)	1. Perform summary statistics of sequences 2. Drop low quality reads 3. Trim low quality bases 4. Crop reads to certain length 5. Filter reads by GC content
NGS QC Toolkit (Patel and Jain, 2012)	Single-end & Paired-end (Illumina, Roche/454)	1. Perform summary statistics of sequences 2. Cut primer, adapter and other artificial sequences 3. Drop low quality reads 4. Crop reads to certain length 5. Trim homopolymer
FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/)	Single-end (Illumina, 454/Roche, Ion Torrent)	1. Perform summary statistics of sequences 2. Cut primer, adapter and other artificial sequences 3. Drop low quality reads 4. Trim low quality bases 5. Crop reads to certain length
ClinQC (Pandey *et al.*, 2016)	Single-end & Paired-end (Illumina, 454/Roche, Ion Torrent, Sanger)	1. Perform summary statistics of sequences 2. Cut primer, adapter and other artificial sequences 3. Drop low quality reads 4. Trim low quality bases
SolexaQA (Cox *et al.*, 2010)	Single-end & Paired-end (Illumina, 454/Roche, Ion Torrent)	1. Perform summary statistics of sequences 2. Drop low quality reads 3. Trim low quality bases

1.11 Figures

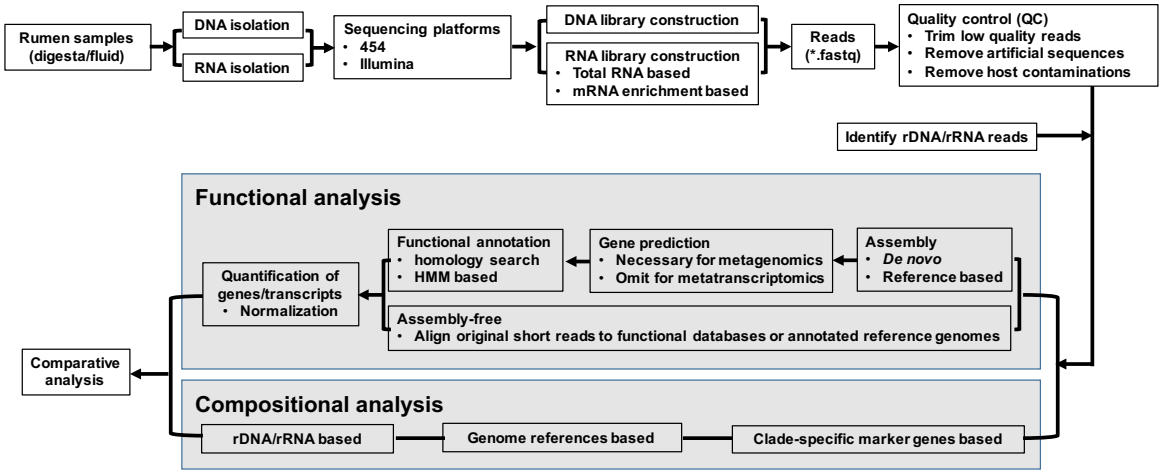


Figure 1.1 A general workflow to conduct rumen metagenomics and metatranscriptomics. To perform the functional analysis of metatranscriptomic data, it is necessary to discard rRNA reads and only include putative mRNA into the analysis.

Chapter 2. Taxonomic assessment of rumen microbiota using total RNA and targeted amplicon sequencing approaches

2.1 Introduction

Microbiota play essential roles in many ecosystems, including the animal gastrointestinal tract, and have attracted much attention in the past decade due to the understanding of their functions in host productivity and health (Holmes *et al.*, 2012; Million *et al.*, 2013; Yeoman and White, 2014). Numerous studies have shown that changes in gastrointestinal microbiota at the taxonomic and/or functional levels are associated with host dysfunction and metabolic diseases (Marchesi *et al.*, 2015; Ojeda *et al.*, 2015), highlighting the importance of studying the interactions that exist between gastrointestinal microbiota and host animals. Therefore, an accurate assessment of the composition and diversity of rumen microbiota is essential to link microbiota changes to host performance under different conditions.

To date, culture-independent molecular-based taxonomic assessment of microbiota has primarily relied on the sequencing of PCR amplicons of targeted microbial genes at the DNA level (DNA Amplicon-seq). Although Amplicon-seq has been widely used, it can be biased due to primer selection (Hong *et al.*, 2009) and/or amplification cycling conditions (Huber *et al.*, 2009). It is also limited in discovering novel microbial phylotypes because the associated primers are designed based on known sequences (Ross *et al.*, 2012; Urich *et al.*, 2008). In addition, to study different groups of microbes within the same microbiota, a wide range of primers is needed

(Kittelmann *et al.*, 2013). Total DNA sequencing (metagenomics) has also been widely used to study microbiota without PCR amplification, and provides information on the presence and absence of phylotypes, but it mainly offers insight in microbial functions through studying microbiota-associated genes. Although a couple of studies have assessed microbial profiles based on 16S rDNA sequences generated in metagenomics datasets (Ellison *et al.*, 2014; Logares *et al.*, 2014), most metagenomic studies rely on parallel DNA Amplicon-seq to characterize microbial communities (Mason and Scott, 2014; Rooks *et al.*, 2014) due to the low fraction of 16S rDNA reads present in metagenomics datasets (Logares *et al.*, 2014). Meanwhile, DNA-based methods do not directly measure the activity of the microbiota because they cannot distinguish the presence of genes that stem from active cells, inactive but alive cells, dead cells, or lysed cells (Gaidos *et al.*, 2011).

To overcome these limitations of DNA-based approaches, recent improvements in RNA sequencing have created a great opportunity to study potentially active microbiota. However, RNA sequencing has mainly been applied to elucidate the functions of microbiota through the mRNA enrichment (de Menezes *et al.*, 2012; Franzosa *et al.*, 2014) and to study active phylotypes through 16S rRNA amplicon sequencing (RNA Amplicon-seq) (Gaidos *et al.*, 2011; Kang *et al.*, 2013). Total RNA sequencing (RNA-seq) has been explored for taxonomic assessment in a number of environments, including soil (Tveit *et al.*, 2014; Urich *et al.*, 2008), hydrothermal vents (Lanzen *et al.*, 2011), and the animal gut (Poulsen *et al.*, 2013; Schwab *et al.*, 2014). However, most of these studies did not compare outcomes between DNA- and RNA-

based methods for the same samples and did not compare RNA-seq vs. Amplicon-seq (DNA/RNA), except Berry *et al.* (2012), who used DNA Amplicon-seq and RNA-seq to study shifts in murine gut microbiota in dextran sodium sulfate (DSS)-induced colitis, and Lanzen *et al.* (2011), who explored microbial communities at both the DNA and RNA levels in the hydrothermal vents. To date, it is not conclusive which method is the most reliable to assess animal gastrointestinal microbial communities because the different outcomes of these methods have not yet been comprehensively compared.

In this study, we compared bacterial and archaeal community profiles in rumen digesta samples using RNA-seq and RNA/DNA Amplicon-seq with standard protocols and a pipeline developed in house. The rumen microbial community is complex and includes bacteria, archaea, protozoa and fungi (Edwards *et al.*, 2004). Although Poulsen *et al.* (2013) used RNA-seq to study rumen microbiota, they mainly focused on methanogens and did not analyze bacteria or compare RNA-seq with Amplicon-seq (DNA and RNA). In this community study, our aim was to gain a better understanding of the differences between the techniques using different genetic materials (DNA vs. RNA) and how they affect interpretation of microbiota-associated data.

2.2 Materials and Methods

2.2.1 Animals and sampling

Rumen digesta samples were collected from five 10-month-old crossbred beef steers, which were raised under feedlot conditions on a high-energy finishing diet, as previously

described (Hernandez-Sanabria *et al.*, 2013) and followed the guidelines of the Canadian Council on Animal Care (Olfert *et al.*, 1993). The animal protocol was approved by the Animal Care and Use Committee of University of Alberta (protocol no. Moore-2006-55). Animals were not starved before the sampling, and were slaughtered before feeding. For each animal, ~3 g of rumen digesta were collected at slaughter and stored in RNA*later* (Ambion, Carlsbad, CA, USA) at -20 °C for further analysis.

2.2.2 Nucleic acid extractions

Total RNA was extracted from rumen digesta using a modified procedure based on the acid guanidinium-phenol-chloroform method (Béra-Maillet *et al.*, 2009; Chomczynski and Sacchi, 1987). Specifically, for ~200 mg of rumen digesta sample, 1.5 ml of TRIzol reagent (Invitrogen, Carlsbad, CA, USA), 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) were used. RNA quality and quantity was determined with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), respectively. RNA samples with the RNA integrity number (RIN) higher than 7.0 were used for downstream analysis. DNA was extracted from 25-30 mg of freeze-dried and ground rumen digesta according to the PCQI method (modified phenol-chloroform with bead beating and QIAquick PCR purification kit) (Henderson *et al.*, 2013; Rius *et al.*, 2012).

2.2.3 RNA library construction and sequencing (RNA-seq)

Total RNA (5 µl of 20 ng/µl) from each sample was used to construct an RNA library following the TruSeq RNA sample Prep v2 LS protocol (Illumina, San Diego, CA, USA), without the mRNA enrichment (rRNA removal) step. The quality and concentration of cDNA fragments (~260 bp) containing artificial sequences (adapters, index sequences, and primers; ~120 bp) and inserted cDNA sequences (~140 bp) were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies) and a Qubit 2.0 fluorometer (Invitrogen), respectively, before sequencing. RNA libraries were paired-end sequenced (2 × 100 bp) using an Illumina HiSeq2000 platform (McGill University and Génome Québec Innovation Centre, QC, Canada).

2.2.4 Amplicon-seq of 16S rRNA/rDNA using pyrosequencing (RNA/DNA Amplicon-seq)

For the DNA Amplicon-seq, partial bacterial and archaeal 16S rRNA genes (the V1-V3 region for bacteria and the V6-V8 region for archaea) were amplified as previously described by Kittelmann *et al.* (2013) and sequenced using 454 GS FLX Titanium chemistry at Eurofins MWG Operon (Ebersberg, Germany). For the RNA Amplicon-seq, total RNA was first reverse-transcribed into cDNA using SuperScript II reverse transcriptase (Invitrogen) with random primers following procedures for first-strand cDNA synthesis. Then, partial 16S rRNA amplicons of bacteria and archaea were generated using the same primers as for the DNA Amplicon-seq and sequenced using a 454 pyrosequencing platform at McGill University and Génome Québec Innovation Centre (Montreal, QC, Canada).

2.2.5 Analysis of the RNA-seq dataset

The sequence data quality was checked using the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The program 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. The filtered reads were then sorted to enrich for 16S rRNA fragments for taxonomic identification and mRNA reads for functional analysis (not reported in this study) using SortMeRNA (version 1.9) (Kopylova *et al.*, 2012) by aligning with the rRNA reference databases SILVA_SSU (release 119), SILVA_LSU (release 119) (Quast *et al.*, 2013), and Rfam 11.0 (Burge *et al.*, 2013). After the 16S rRNA sequences were enriched, downstream analyses were performed using the mothur program (version 1.31.2) (Schloss *et al.*, 2009) according to the procedures (http://www.mothur.org/wiki/MiSeq_SOP) described by Kozich *et al.* (2013), with modifications. For taxonomic identification, regionally enriched reference datasets were built for bacteria and archaea (mothur command: pcr.seqs). Specifically, sequences belonging to the V1-V3 region (mean length: 466 bp) were extracted from the aligned Greengenes 16S rRNA gene database (version gg_13_5_99 accessed May 2013) (DeSantis *et al.*, 2006) for bacteria. For archaea, sequences belonging to the V6-V8 region (mean length: 456 bp) were extracted from the aligned rumen-specific archaeal 16S rRNA gene database derived from a previous study (Janssen and Kirs, 2008). The starting and ending positions of the targeted regions were located based on the alignment of Amplicon-seq reads to

the references databases (**Figure 2.1**) because these amplicons were generated using designed primers for known regions.

The overall RNA-seq data analysis pipeline is illustrated in **Figure 2.1**. Briefly, sorted paired-end reads belonging to bacterial and archaeal 16S rRNA were joined (mothur command: make.contigs) to increase the length by combining the forward and reverse sequences. Joined sequences (mean length: 140 bp) with ambiguous bases or longer than 200 bp were discarded (mothur command: screen.seqs) to remove sequences without overlapped regions. Identical sequences were then binned to generate a set of unique sequences to facilitate the counting of their frequencies in each sample (mothur commands: unique.seqs and count.seqs). Next, the bacterial and archaeal 16S rRNA sequences in the sample datasets were aligned to the regionally enriched bacterial (the V1-V3 region) and archaeal (the V6-V8 region) references (see above), respectively (mothur command: align.seqs with default option). Sequences for which less than 50 bp aligned to the reference datasets were culled (mothur command: screen.seqs). After alignment filtering (mothur command: filter.seqs), combining the identical sequences, and counting the frequencies (mothur command: unique.seqs), pre-clustering was performed to decrease the complexity of our sample datasets by clustering highly similar sequences with one nucleotide identity difference (mothur command: pre.cluster). UCHIME (Edgar *et al.*, 2011) in *de novo* mode and with default settings was applied to identify and remove chimeras (mothur command: chimera.uchime & remove.seqs). Finally, chimera-depleted sequences were subject to taxonomic assignment to different phylotypes using a naive Bayesian algorithm (Wang *et al.*, 2007) with a

minimum confidence of 0.8 (mothur command: classify.seqs), and then the taxonomic rank and the relative abundance for each phylotype was calculated (mothur commands: phylotype, make.shared, classify.otu, and get.relabund). To make alpha-diversity estimators comparable among samples and different methods, all samples were standardized to the same number of sequences (the smallest sampling size) by randomly selecting sequences from the chimera-depleted datasets (mothur command: sub.sample). Alpha-diversity analysis was conducted at the bacterial family level and archaeal species level. The Good's coverage, the Shannon index, the inverse Simpson index, the number of observed phylotypes, and the Chao estimator were calculated based on the normalized samples (mothur command: summary.single).

2.2.6 Analysis of RNA/DNA Amplicon-seq datasets

The procedures were similar to those described for the RNA-seq dataset (**Figure 2.1**). Briefly, after trimming primers and screening homopolymer runs (maximum length: 6), only sequences over 200 bp in length with an average quality score over 25 and with less than 6 ambiguous bases were included in the analysis. This step was performed using mothur (version 1.31.2) (Schloss *et al.*, 2009) with the command t*rim.seqs*. After clustering similar sequences, the chimeras were checked and discarded from reads with good quality. The chimera-depleted reads were used for taxonomic identification and to calculate the relative abundance of each phylotype. The alpha-diversity was analyzed using the standardized chimera-depleted sequences according to the lowest number of reads.

2.2.7 Validation of bacterial relative abundance using qRT-PCR and qPCR

Quantitative reverse transcription PCR (qRT-PCR) and quantitative PCR (qPCR) were further performed to validate the relative abundance data obtained from RNA-seq, RNA Amplicon-seq, and DNA Amplicon-seq. Three primer pairs (**Table 2.5**) were used to enumerate the total bacteria, *Bacteroidetes,* and *Gammaproteobacteria* in each rumen sample. Standard curves were constructed using serial dilutions of plasmid DNA from clones identified as *Butyrivibrio hungatei* (for total bacteria, using an initial concentration of 8.50×10^8 mol/μl), *Prevotella sp.* (for *Bacteroidetes*, using an initial concentration of 2.89×10^8 mol/μl) and *Tolumonas auensis* (for *Gammaproteobacteria*, using an initial concentration of 2.68×10^8 mol/μl). The copy numbers for each standard curve were calculated as described previously (Li *et al.*, 2009b). For qRT-PCR, cDNA was first reverse-transcribed from 20 ng of total RNA using iScrtpt reverse transcription supermix for RT-qPCR (Bio-Rad, Hercules, CA, USA) and then diluted 5 times. One microliter of diluted cDNA was subjected to a qRT-PCR reaction using SYBR Green chemistry (Fast SYBR Green Master Mix; Applied Biosystems) in a StepOnePlus Real-Time PCR System (Applied Biosystems). The relative abundances of *Bacteroidetes* and *Gammaproteobacteria* compared to total bacteria were calculated according to the following equation: relative abundance = Q_{Target} / Q_{U2}, where Q_{Target} was the quantity of each target, and Q_{U2} was the total bacteria quantity. Concurrently, we performed qPCR using total DNA (1 μl × 10 ng/μl total DNA per reaction) and followed the same procedures mentioned above to verify DNA Amplicon-seq results.

2.2.8 Statistical analysis

In this study, only taxa with a relative abundance > 0.1% in at least two samples within the RNA-seq, RNA Amplicon-seq, and DNA Amplicon-seq datasets were defined as detectable and subjected to downstream statistical analysis. Statistical summaries (mean and SEM) of the detected taxa, and Pearson's correlation analysis for qRT-PCR/qPCR validation were all performed using R 3.1.2 (R Core Team, 2014). Principal coordinate analysis (PCoA) of the microbial profiles generated from the different datasets was conducted based on the Bray-Curtis dissimilarity matrix. The microbial relative abundance was arcsine-square-root transformed (Franzosa *et al.*, 2014), and then Repeated Measures ANOVA was performed to compare the differences among three datasets. *P* values were adjusted into FDR using Benjamini-Hochberg method (Benjamini and Hochberg, 1995), and a threshold of FDR < 0.15 was applied to determine the significance (Korpela *et al.*, 2016). Co-occurrence analysis was performed for the bacterial families and archaeal taxa detected in both the RNA-seq and DNA Amplicon-seq datasets (with relative abundance > 0.1% in all five samples) based on Spearman's rank correlation (Barberan *et al.*, 2012). An association network was constructed using CoNet (Faust *et al.*, 2012) and displayed using Cytoscape 3.2.1 (Faust *et al.*, 2012; Shannon *et al.*, 2003). Effective alpha-diversity estimators (Jost, 2007) were compared among the RNA-seq, RNA Amplicon-seq, and DNA Amplicon-seq datasets using paired Wilcoxon signed rank test.

2.2.9 Data submission

RNA-seq and RNA Amplicon-seq datasets were submitted into the NCBI Sequence Read Archive (SRA) under the accession number PRJNA275012, and DNA Amplicon-seq sequences were also placed in the NCBI SRA under the accession number PRJNA273417.

2.3 Results and Discussion

2.3.1 Analyzing rumen microbiota using RNA-seq and RNA/DNA Amplicon-seq

This study assessed active rumen microbial communities using RNA-seq and is the first study to compare RNA-seq outcomes with the well-accepted Amplicon-seq methods to evaluate rumen microbiota. It has been reported that rRNA levels directly relate to the protein synthesis potential of microorganisms (Blazewicz *et al.*, 2013) and are correlated with activity (Bremer and Dennis, 2008; Poulsen *et al.*, 1993). rRNA abundance data obtained from total RNA sequencing could potentially be used as one of the indices to taxonomically assess potentially active microbes within a sample. To explore the possibility of taxonomic profiling using total RNA-seq, we first enriched 16S rRNA sequences from an RNA-seq dataset (**Figure 2.1**). In total, an average of 38,496,238 ± 2,037,011 (mean ± SEM) reads per sample (192,481,188 reads in total) were obtained after quality control filtration. Among them, 92.9 ± 1.1% belonged to small and large subunit rRNA, with 13.7 ± 5.6% and 0.2 ± 0.0% being bacterial and archaeal 16S rRNA, respectively (**Table 2.1**). It is notable that a large fraction of rRNA was classified as eukaryotic 18S (22.1 ± 5.9%) and 28S (32.2 ± 8.2%) rRNA (**Table 2.1**). Although these reads were not

analyzed in the current study, the high number of these sequences indicates the possibility of assessing rumen eukaryotic microbiota using RNA-seq in future studies. After combining paired-end reads and removing non-overlapping sections, 10,782,833 bacterial and 152,585 archaeal joint 16S rRNA sequences were obtained.

In this study, we included regionally enriched 16S rRNA gene reference datasets for taxonomic analysis rather than aligning the sequences to the full-length 16S rRNA gene database directly. Because the length of Illumina RNA-seq reads is short, 16S rRNA sequences could be randomly aligned to different regions of the 16S rRNA gene. It is known that different hypervariable regions of the 16S rRNA gene can affect diversity estimation and taxonomic classification (Logares *et al.*, 2014). If these short 16S rRNA sequences are directly mapped to full-length 16S rRNA gene references for taxonomic analysis, as has been performed in previous studies (Tveit *et al.*, 2014; Urich *et al.*, 2008), it could lead to a mixed taxonomic profile as well as an overestimation of diversity. To avoid such potential bias, two regionally enriched reference datasets were generated from the Greengenes 16S rRNA gene database (version gg_13_5_99 accessed May 2013) (the V1-V3 region for bacteria) and the rumen-specific archaeal 16S rRNA database (Janssen and Kirs, 2008) (the V6-V8 region for archaea) (see details in the *Materials and Methods* section). These regions were chosen because the V2-V3 region is the most efficient region for assessing bacterial community (Chakravorty *et al.*, 2007), while the V6-V8 region is the most efficient region for identifying archaea and estimating the archaeal community diversity (Snelling *et al.*, 2014). After identifying sequences belonging to the bacterial V1-V3 region and

the archaeal V6-V8 region, sequences aligned with more than 50 bp of the reference datasets were then subjected to chimeric sequence detection (2,423,139 bacterial and 25,451 archaeal). After the removal of 9,353 bacterial sequences (0.4%) and 139 archaeal sequences (0.6%) through chimera checking, 2,413,786 of the V1-V3 region-enriched bacterial sequences (mean length: 124 bp) and 25,312 of the V6-V8 region-enriched archaeal sequences (mean length: 133 bp) were subject to further taxonomic analysis (**Table 2.2**).

RNA and DNA Amplicon-seq of the same rumen samples generated 37,105 (7,421 ± 506; mean ± SEM; RNA Amplicon-Seq dataset) and 31,031 (6,206 ± 645; DNA Amplicon-Seq dataset) bacterial reads, respectively, as well as 8,303 (1,661 ± 20 for RNA Amplicon-Seq) and 6,663 (1,333 ± 95 for DNA Amplicon-Seq) archaeal reads, respectively (**Table 2.1**). From these two datasets, 6,461 / 5,505 bacterial reads (17.4% / 17.7%) and 562 / 662 archaeal reads (6.8% / 9.3%) were detected and removed as chimeric sequences. In total, 30,644 / 25,526 bacterial reads (mean length: 476 / 487 bp) and 7,741 / 6,041 archaeal reads (mean length: 451 / 462 bp) were used for taxonomic identification and quantification (**Table 2.2**).

In this study, we used the *de novo* (database-independent) mode to determine chimeras rather than reference-based chimera detection. This is because the existing chimera reference databases only contain sequences from cultured organisms (Haas *et al.*, 2011) and are not suitable for real samples that contain uncultured bacteria and archaea. Notably, higher percentages of chimeric sequences (17.4% / 17.7% of the bacterial and 6.8% / 9.3% of the archaeal sequences) were removed from the RNA/DNA Amplicon-seq datasets than from the RNA-seq dataset, which

had only 0.4% of their bacterial and 0.6% of their archaeal sequences removed due to the presence of chimeras. In Amplicon-Seq datasets, chimeras are produced during PCR amplification, and they can lead to biased estimation of the diversity and/or the identification of differences between microbial communities (Edgar *et al.*, 2011), while in RNA-seq datasets, chimeras might stem from the cDNA synthesis and/or fragment enrichment procedures used during the library construction. Our results indicate that RNA-seq was less affected by chimera formation than was Amplicon-seq.

2.3.2 Microbial taxa detected from RNA-seq and RNA/DNA Amplicon-seq

From the RNA-seq dataset, 94.6% and 86.9% of the bacterial V1-V3 region-enriched sequences were classified at the phylum and family level, respectively (**Table 2.2**). Due to their short sequence lengths, 86.2% of bacterial sequences from the RNA-seq dataset could not be classified at the genus level. Thus, only bacterial taxa at the phylum and family levels were retained for further analysis. For archaea, 98.0% of the V6-V8 region-enriched sequences were classified in a mixed taxonomic rank scheme (**Table 2.2**). We classified archaeal sequences at different taxonomic levels because most of the predominant archaeal phylotypes (such as *Methanobrevibacter ruminantium* and *Methanobrevibacter gottschalkii*) have well-studied 16S rRNA genes for use as references (Janssen and Kirs, 2008), and even short reads could be classified at the species level. However, for the poorly studied groups, such as *Methanomassiliicoccales*, reads could be only classified at the order level based on this rumen-

specific archaeal database (Janssen and Kirs, 2008). From RNA/DNA Amplicon-seq datasets, 98.2% / 98.8%, 86.8% / 87.1% and 98.6% / 99.7% of the total reads were assigned at the bacterial phylum, bacterial family and archaeal mixed taxon levels, respectively (**Table 2.2**).

The bacterial and archaeal taxa detected in the three datasets were generally similar, with a total of eleven bacterial phyla, twenty-one bacterial families and six archaeal taxa identified. Of these, seven bacterial phyla, fifteen bacterial families and five archaeal taxa were commonly detected across the three datasets (**Figure 2.2**). Notably, there were unique bacterial and archaeal taxa identified in each dataset (**Figure 2.2d**). Firstly, two bacterial families and one archaeal taxon (*Desulfovibrionaceae, Sphaerochaetaceae*, and *Methanobrevibacter woesei*) were detected only in the RNA-seq dataset and not in the RNA/DNA Amplicon-seq datasets. Henderson *et al.* (2015) also reported the absence of *Desulfovibrionaceae* and *Methanobrevibacter woesei*, and the low abundance (≤ 0.1%) of *Sphaerochaetaceae* in the rumen digesta when using DNA Amplicon-seq with the same PCR primers, suggesting that the amplicon-based approach and/or the primers used may have masked the detection of these taxa in the rumen. In addition, lower sequencing depth of RNA/DNA Amplicon-seq could also lead to the missing detection of these taxa. As shown in **Table 2.2**, the number of bacterial and archaeal reads from the RNA-seq dataset was about 80-90 times higher and 3-4 times higher than that from RNA/DNA Amplicon-seq datasets, respectively. And thus increasing sequencing depth could probably enhance the detection of these taxa in the Amplicon-seq datasets. Secondly, two bacterial phyla (*Elusimicrobia* and *Verrucomicrobia*) and one bacterial family (*Elusimicrobiaceae*) were

detected only in the RNA-based datasets (RNA-seq and RNA Amplicon-seq) and not in the DNA Amplicon-seq dataset. Our results suggest that these two phyla may be more active in the rumen, and they may be underestimated based on the DNA Amplicon-seq dataset. Moreover, the absence of *Elusimicrobia* and *Verrucomicrobia* in the DNA Amplicon-seq dataset may be due to the unsuccessful isolation of their DNA, and it has been reported that various DNA extraction methods could impact the taxonomic outcomes of rumen microbiota assessments (Henderson *et al.*, 2013). Finally, the bacterial family *Streptococcaceae* was detected only in the DNA Amplicon-seq dataset with a low relative abundance of 0.1 ± 0.1%. Previous studies on the bacterial profiles of rumen digesta from the same cattle used in this study (Hernandez-Sanabria *et al.*, 2012) and different cattle (Petri *et al.*, 2013; Xia *et al.*, 2015) have also reported the absence of *Streptococcaceae*. In a recent study based on a large number (n = 742) of rumen and foregut digesta samples and DNA-based Amplicon-seq, *Streptococcaceae* also showed low prevalence in all animals (Henderson *et al.*, 2015). These suggest that *Streptococcaceae* may have low cellular abundance and even lower activities in samples assessed in the current study.

2.3.3 Estimated microbial relative abundance from RNA-seq and RNA/DNA Amplicon-seq

Principal coordinate analysis (PCoA) of the relative abundances of commonly detected bacterial phyla, bacterial families, and archaeal taxa revealed dissimilarities in microbial profiles between RNA- and DNA-based approaches (**Figure 2.3**). For each animal, the RNA-seq and RNA Amplicon-seq generated similar rumen bacterial profiles (at both the phylum and family

levels), which generally separated with that assessed using DNA Amplicon-seq (**Figure 2.3a** and **2.3b**). However, for each animal, three datasets displayed generally similar archaeal profiles (**Figure 2.3c**); the RNA-based assessment outcomes of L52 and L59 were distinct from their DNA-based profiles because these two samples had a high abundance of *Methanomassiliicoccales* in the RNA-based datasets (**Figure 2.2c**). Among the shared taxa, four bacterial phyla (*Bacteroidetes*, *Lentisphaerae*, *Proteobacteria*, and *Synergistetes*), eight bacterial families (*Dethiosulfovibrionaceae*, *Lactobacillaceae*, *Mogibacteriaceae*, *Paraprevotellaceae, Prevotellaceae,* S24-7, *Succinivibrionaceae*, and *Victivallaceae*), and one archaeal taxon (*Methanomassiliicoccales*) had significantly different relative abundances among the three datasets (FDR < 0.15; **Table 2.3**). Lanzen *et al.* (2011) reported that the dominant taxa within hydrothermal vent field microbiota showed similar outcomes on the RNA and DNA levels based on their Amplicon-seq results, which is not consistent with our findings and is probably an extremely environment-specific case. In addition, such discrepancy may also be due to different targeted amplicon regions (the V5-V6 region in Lanzen *et al.* (2011) vs. the V1-V3 / V6-V8 regions in the current study) as well as different reference databases used (the Silva SSURef in Lanzen *et al.* (2011) vs. the regionally enriched Greengenes/rumen-specific archaeal databases in our study). Therefore, different methods and strategies should be carefully considered for samples from various environmental conditions.

The dominant bacterial phyla detected by DNA Amplicon-seq were *Bacteroidetes* (50.3 ± 8.7%; mean ± SEM), *Firmicutes* (29.4 ± 6.2%), and *Proteobacteria* (14.3 ± 8.5%), which is

consistent with previous studies using DNA-based methods. For example, *Bacteroidetes* (range, 8.0 to 60.1%) and *Firmicutes* (range, 33.6 to 85.0%) were reported as the most abundant phyla, and *Proteobacteria* was commonly detected but less abundant (range, 0.6 to 20.1%) in the rumen (Jami and Mizrahi, 2012; Kim and Yu, 2014; Petri *et al.*, 2013). However, the predominant bacterial phylum detected by RNA-based approaches (RNA-seq and RNA Amplicon-seq) was *Proteobacteria* (47.6 ± 14.2% and 46.3 ± 14.3%, respectively), followed by *Bacteroidetes* (23.3 ± 7.9% and 22.7 ± 8.1%), and *Firmicutes* (16.2 ± 4.5% and 19.2 ± 6.0%). The higher proportion of *Proteobacteria* in the RNA-based datasets confirmed similar findings by Kang *et al.* (2013) and Kang *et al.* (2009), who applied RNA amplicon-based sequencing and rRNA-based clone libraries, respectively, to study the rumen microbiota. At the bacterial family level, the most abundant bacterial family was *Succinivibrionaceae* (belonging to the phylum *Proteobacteria*) in the RNA-seq (45.6 ± 14.0%) and RNA Amplicon-seq (45.1 ± 14.4%) datasets, while it was *Prevotellaceae* (belonging to the *Bacteroidetes*) in the DNA Amplicon-seq dataset (38.7±8.6%). *Succinivibrionaceae* was an abundant family at the DNA level when ruminants were fed high-energy diets (Henderson *et al.*, 2015; Hernandez-Sanabria *et al.*, 2012; Petri *et al.*, 2013), and the significance of *Succinivibrionaceae* may be underestimated using DNA Amplicon-seq. The predominance of *Prevotellaceae* detected in the DNA Amplicon-seq dataset is similar to that observed in previously studies using DNA-based approaches (Henderson *et al.*, 2015; Kittelmann *et al.*, 2013; Petri *et al.*, 2013). However, its abundance, as estimated in the RNA-seq (16.4 ± 5.7%) and RNA Amplicon-seq (17.9 ± 7.2%) datasets, was significantly lower (FDR < 0.15). The

family *Lactobacillaceae*, belonging to the phylum *Firmicutes*, also had a lower abundance in the RNA-based datasets (0.3 ± 0.1% using RNA-seq and 0.2 ± 0.1% using RNA Amplicon-seq) than in the DNA Amplicon-seq dataset (8.1 ± 5.0%, FDR < 0.15). These two families probably had higher cellular abundance but relatively lower activities in the bovine rumen.

From the RNA/DNA Amplicon-seq datasets, *Methanobrevibacter gottschalkii* (28.2±9.7% / 37.1±10.6%), *Methanobrevibacter ruminantium* (37.1 ± 12.4% / 36.4 ± 13.1%), and *Methanomassiliicoccales* (30.9 ± 10.0% / 15.7 ± 2.0%) were dominant but with different rankings. The DNA Amplicon-seq outcomes are generally consistent with previous studies that used the same approaches (Henderson *et al.*, 2015; Kittelmann *et al.*, 2013). The archaeal taxon *Methanomassiliicoccales*, which has been previously referred to as Rumen Cluster C or *Thermoplasmatales* (Gaci *et al.*, 2014; Janssen and Kirs, 2008; Poulsen *et al.*, 2013), was predominant in the RNA-seq dataset (38.8 ± 11.9%), followed by *Methanobrevibacter ruminantium* (30.2 ± 10.4%) and *Methanobrevibacter gottschalkii* (22.4 ± 7.4%). The high proportion of *Methanomassiliicoccales* from the RNA-based datasets supports the hypothesis that they are more active in the rumen, as many studies have suggested (Janssen and Kirs, 2008; Jeyanathan *et al.*, 2011; Ohene-Adjei *et al.*, 2007; Williams *et al.*, 2009; Wright *et al.*, 2007).

2.3.4 qRT-PCR and qPCR validation of estimated relative abundance among datasets

qRT-PCR and qPCR were performed to estimate the relative abundances of two predominant phyla (*Bacteroidetes* and *Proteobacteria*) among all three datasets.

Gammaproteobacteria was selected to represent *Proteobacteria* because 95.8%, 98.0% and 96.5% of the *Proteobacteria* reads from the three datasets belonged to the class *Gammaproteobacteria*. The qRT-PCR results for *Bacteroidetes* were in agreement with the RNA-seq (Pearson's correlation coefficient [r] = 0.97, $P < 0.05$) and the RNA Amplicon-seq (r = 0.97, $P < 0.05$) results, and the qPCR results for *Bacteroidetes* were also consistent with the DNA Amplicon-seq results (r = 0.88, P = 0.05) **(Figure 2.4)**. The relative abundance of *Gammaproteobacteria* estimated using qRT-PCR was correlated with that from the RNA-seq dataset (r = 0.97, $P < 0.05$) and the RNA Amplicon-seq dataset (r = 0.99, $P < 0.05$), and there was also a high degree of correlation between qPCR and DNA Amplicon-seq (r = 0.99, $P < 0.05$) for *Gammaproteobacteria*. The overall consistent trends between the RNA-based approaches and qRT-PCR (as well as between DNA Amplicon-seq and qPCR) confirm the different relative abundance detected in the three datasets.

2.3.5 Alpha-diversity estimators in the RNA-seq and RNA/DNA Amplicon-seq datasets

In the present study, alpha-diversity indices were estimated based on the observed phylotypes at the family level for bacteria and at the species level for archaea. To avoid potential differences caused by different sequencing depths among the three datasets, the samples were randomly normalized according to the lowest number of reads (3,476 bacterial sequences and 1,074 archaeal sequences per sample after normalization). The values of Good's coverage were all above 99% for the bacterial and archaeal data from the three datasets, indicating that the

numbers of reads after normalization were sufficient to represent the microbial communities. The Shannon index and the inverse Simpson index were not significantly different ($P > 0.1$, the paired Wilcoxon signed rank test) among the three datasets (**Table 2.4**). The number of observed phylotypes and the Chao estimator tended to be higher in the RNA-seq dataset than in the RNA/DNA Amplicon-seq datasets for bacteria and archaea ($P < 0.1$, the paired Wilcoxon signed rank test; **Table 2.4**), which was further confirmed using rarefaction analysis (**Figure 2.5**). These results suggest that more microbial taxa could be detected using RNA-seq than using RNA/DNA Amplicon-seq. In the Amplicon-seq datasets, some phylotypes were probably overlooked due to the bias of primers and/or amplification conditions during the PCR process, which may explain the difference in richness among the three datasets.

2.3.6 Co-occurrence analysis of abundant microbial taxa detected by RNA-seq and DNA Amplicon-seq

The relationships that exist among microbial taxa could be a determining factor for microbial community composition (Prosser *et al.*, 2007). To explore the relationships among different taxa in our samples, Spearman's rank correlation was used to identify the co-occurrence patterns of different microbial groups in both the RNA-seq and DNA Amplicon-seq datasets (**Figure 2.6**). The RNA Amplicon-seq dataset was not included due to its similar outcomes to the RNA-seq dataset, with high correlation for all samples (Spearman's rank correlation coefficient [ρ] = 0.88 - 0.98, $P < 0.0001$) (**Figure 2.2** and **2.3**). As shown in **Figure 2.6**, the microbial taxa

identified using RNA-seq were more closely correlated than those identified using DNA Amplicon-seq. There may be stronger interactions among microbes within a microbiota at the transcriptional level than that at the genomic level, which has been suggested recently by İnceoğlu *et al.* (2015). Previous studies have revealed associations among microbes (*Ruminococcaceae* and *Mbb. gottschalkii, Succinivibrionaceae* and *Methanomassiliicoccales*, *Mbb. gottschalkii* and *Mbb. ruminantium*, and *Methanomassiliicoccales* and *Methanosphaera*, etc) in rumen using DNA Amplicon-seq (Henderson *et al.*, 2015; Kittelmann *et al.*, 2013). In this study, we also detected a negative relationship between *Mbb. gottschalkii* and *Mbb. ruminantium* (ρ = -0.9, $P < 0.1$; **Figure 2.6**) in the DNA Amplicon-seq dataset. In the RNA-seq dataset, we confirmed that *Succinivibrionaceae* was positively correlated with *Methanomassiliicoccales* (ρ = 0.9, $P < 0.1$) (**Figure 2.6**). The bacterial families *Lachnospiraceae, Mogibacteriaceae, Prevotellaceae*, *Ruminococcaceae,* and *Spirochaetaceae* were positively correlated with each other ($P < 0.1$), and all of them were negatively correlated with *Succinivibrionaceae* and *Methanomassiliicoccales* ($P < 0.1$) in the RNA-seq dataset (**Figure 2.6b**). However, the negative correlations between *Succinivibrionaceae* and the other 4 bacterial families, and between *Mbb. gottschalkii* and *Mbb. ruminantium*, are possibly because they displayed the arithmetic replacement effect (shifts in abundance of predominant phylotypes will have effects on others when analyzing proportion data) as suggested by Henderson *et al.* (2015). Meanwhile, only two taxa, *Succinivibrionaceae* and *Mbb. ruminantium*, showed significantly positive correlations between the RNA-seq and DNA Amplicon-seq datasets (ρ = 1.0, $P < 0.05$ and ρ = 0.9, $P < 0.1$,

respectively). No other taxa exhibited strong consistency between the RNA-seq and DNA Amplicon-seq datasets, indicating that cellular abundance did not correspond to the activities of most rumen taxa.

2.3.7 Methodological caveats of this study

This study has limitations that should be taken into account. Firstly, the sampling timing may have more of an effect on RNA-based than DNA-based approaches of profiling microbial communities. If assessing the activities of rumen microbiota is the main study objective, the sampling timing should be carefully considered. In this study, rumen digesta samples were collected before the feeding, which probably resulted in different RNA profiles than in digesta samples collected after feeding. However, the same rumen digesta sample was used for RNA and DNA extraction in our study, so the detected differences between RNA- and DNA-based analyses are valid and not biased due to different sampling times. Second, during RNA isolation processes, RNA yield may differ according to extraction method (such as between physical, mechanic, enzymatic, and chemical methods) (Stark *et al.*, 2014). Meanwhile, not all microbes can be lysed with equal efficiency, and notably, RNA yields from Gram-positive bacteria are generally lower than those from Gram-negative bacteria (Stark *et al.*, 2014). For instance, members of *Proteobacteria* and *Bacteroidetes* are Gram-negative, while most *Firmicutes* members in the rumen are Gram-positive, which may explain the higher abundance of *Proteobacteria* and *Bacteroidetes,* and the lower abundance of *Firmicutes* in the RNA-seq

dataset. Moreover, significant differences in microbial community structures were also demonstrated to correspond to different DNA extraction methods in a report by (Henderson *et al.*, 2013). Third, because methods (Griffiths *et al.*, 2000; Leininger *et al.*, 2006) for co-extraction of RNA and DNA could not generate the high quality RNA for RNA-seq in our samples (RNA integrity number < 3.0), the RNA and DNA extraction were conducted separately using two independent protocols to ensure the high quality RNA and DNA in the current study, which could potentially lead to differences between RNA- and DNA-based methods. Furthermore, the RNA was transcribed to cDNA using the random primers before making the RNA-seq and RNA Amplicon-seq libraries, while the DNA Amplicon-seq was performed using regional specific primers to amplify DNA template directly, which could also contribute to differences among three datasets. Fourth, because rRNA content per cell varies between different microbial phylotypes (Medlin and Simon, 1998; Sievert *et al.*, 2000), such intrinsic differences in rRNA content could also influence the relative abundance determined using RNA-seq. Future experiments to globally compare the rRNA content per cell among different microbial phylotypes in the rumen microbiota and normalize rRNA concentrations from different phylotypes can improve the accuracy of microbial community profiling using RNA-seq. Fifth, the quantification of total and/or species-specific rRNA is a valid and well-accepted approach to estimate the microbial activity, which has been applied in more than 100 studies (Blazewicz *et al.*, 2013). However, the use of rRNA as an indicator of specific microbial functional activity in complex environmental samples still need to be further validated by correlating them with the mRNA

information within the same metatranscriptomic datasets. Sixth, in the current study, the RNA-seq dataset was generated on an Illumina HiSeq2000 platform, whereas RNA/DNA Amplicon-seq was performed using a 454 pyrosequencing platform. It has been demonstrated that there are different features between these two platforms, such as read length, accuracy, and throughput (Liu *et al.*, 2012). However, previous studies have revealed the consistency of microbial community profiles generated across sequencing platforms (Caporaso *et al.*, 2012; Nelson *et al.*, 2014), but differences in sequencing depth could have an impact on the detection of low-abundant taxa.

2.4 Conclusion

A comparison of the microbial profiles generated from RNA-seq and RNA/DNA Amplicon-seq revealed the generation of different taxonomic profiles of the same rumen microbiota between these methods, and thus their results could not be simply combined. The RNA-based methods could more robustly detect microbial phylotypes with potentially metabolic activities in the rumen and also detect more interactions among these phylotypes than DNA Amplicon-seq. In addition, compared to RNA/DNA Amplicon-seq, the RNA-seq approach showed more diversity and could detect more bacterial and archaeal phylotypes in the rumen. Although the RNA-seq approach has the advantage of simultaneously identifying and quantifying active microorganisms within a microbiota, the data are not conclusive on which method is the best for analyzing animal gastrointestinal microbiota due to the different technologies and

constraints of DNA vs. RNA and due to differences in nucleotide extraction, sequencing, and analysis protocols. Nevertheless, this is the first study to compare RNA-seq and RNA/DNA Amplicon-seq for the taxonomic assessment of rumen microbiota, and differences among these methods should be carefully considered to accurately assess gastrointestinal microbiota in future studies.

2.5 References

Barberan A, Bates ST, Casamayor EO, Fierer N (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *Isme j* **6:** 343-351.

Benjamini Y, 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:** 289-300.

Béra-Maillet C, Mosoni P, Kwasiborski A, Suau F, Ribot Y, 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. *J Microbiol Methods* **77:** 8-16.

Berry D, Schwab C, Milinovich G, Reichert J, Ben Mahfoudh K, Decker T *et al* (2012). Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis. *Isme j* **6:** 2091-2106.

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *Isme j* **7:** 2061-2068.

Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114-2120.

Bremer H, Dennis PP (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* **3**.

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP *et al* (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* **41:** D226-D232.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N *et al* (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *Isme j* **6:** 1621-1624.

Chakravorty S, Helb D, Burday M, Connell N, Alland D (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* **69:** 330-339.

Chomczynski P, Sacchi N (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162:** 156-159.

de Menezes A, Clipson N, Doyle E (2012). Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. *Environ Microbiol* **14:** 2577-2588.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069-5072.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194-2200.

Edwards JE, McEwan NR, Travis AJ, Wallace RJ (2004). 16S rDNA library-based analysis of ruminal bacterial diversity. *Antonie van Leeuwenhoek* **86:** 263-281.

Ellison MJ, Conant GC, Cockrum RR, Austin KJ, Truong H, Becchi M *et al* (2014). Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. *DNA Res* **21:** 115-125.

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J *et al* (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* **8:** e1002606.

Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM *et al* (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* **111:** E2329-2338.

Gaci N, Borrel G, Tottey W, O'Toole PW, Brugere JF (2014). Archaea and the human gut: new beginning of an old story. *World J Gastroenterol* **20:** 16062-16078.

Gaidos E, Rusch A, Ilardo M (2011). Ribosomal tag pyrosequencing of DNA and RNA from benthic coral reef microbiota: community spatial structure, rare members and nitrogen-cycling guilds. *Environ Microbiol* **13:** 1138-1152.

Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ (2000). Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol* **66:** 5488-5491.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G *et al* (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21:** 494-504.

Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, Noel SJ *et al* (2013). Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One* **8:** e74787.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH (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.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Durunna ON, Moore SS, Guan LL (2012). Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. *Appl Environ Microbiol* **78:** 1203-1214.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Zhou M, Moore SS, Guan LL (2013). Influence of sire breed on the interplay among rumen microbial populations inhabiting the rumen liquid of the progeny in beef cattle. *PLoS One* **8:** e58461.

Holmes E, Li JV, Marchesi JR, Nicholson JK (2012). Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab* **16:** 559-564.

Hong S, Bunge J, Leslin C, Jeon S, Epstein SS (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *Isme j* **3:** 1365-1373.

Huber JA, Morrison HG, Huse SM, Neal PR, Sogin ML, Mark Welch DB (2009). Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ Microbiol* **11:** 1292-1302.

İnceoğlu Ö, Llirós M, Crowe SA, García-Armisen T, Morana C, Darchambeau F *et al* (2015). Vertical Distribution of Functional Potential and Active Microbial Communities in Meromictic Lake Kivu. *Microb Ecol*.

Jami E, Mizrahi I (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS One* **7:** e33306.

Janssen PH, Kirs M (2008). Structure of the archaeal community of the rumen. *Appl Environ Microbiol* **74:** 3619-3625.

Jeyanathan J, Kirs M, Ronimus RS, Hoskin SO, Janssen PH (2011). Methanogen community structure in the rumens of farmed sheep, cattle and red deer fed different diets. *FEMS Microbiol Ecol* **76:** 311-326.

Jost L (2007). Partitioning diversity into independent alpha and beta components. *Ecology* **88:** 2427-2439.

Kang S, Denman SE, Morrison M, Yu Z, McSweeney CS (2009). An efficient RNA extraction method for estimating gut microbial diversity by polymerase chain reaction. *Curr Microbiol* **58:** 464-471.

Kang SH, Evans P, Morrison M, McSweeney C (2013). Identification of metabolically active proteobacterial and archaeal communities in the rumen by DNA- and RNA-derived 16S rRNA gene. *J Appl Microbiol* **115:** 644-653.

Kim M, Yu Z (2014). Variations in 16S rRNA-based microbiome profiling between pyrosequencing runs and between pyrosequencing facilities. *J Microbiol* **52:** 355-365.

Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI *et al* (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS One* **8:** e47879.

Kopylova E, Noe L, Touzet H (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28:** 3211-3217.

Korpela K, Salonen A, Virta LJ, Kekkonen RA, Forslund K, Bork P *et al* (2016). Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nat Commun* **7:** 10410.

Kozich J, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl Environ Microbiol* **79:** 5112-5120.

Lanzen A, Jorgensen SL, Bengtsson MM, Jonassen I, Ovreas L, Urich T (2011). Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiol Ecol* **77:** 577-589.

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW *et al* (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442:** 806-809.

Li M, Penner GB, Hernandez-Sanabria E, Oba M, Guan LL (2009). Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen. *J Appl Microbiol* **107:** 1924-1934.

Liu L, Li Y, Li S, Hu N, He Y, Pong R *et al* (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012:** 251364.

Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento H *et al* (2014). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* **16:** 2659-2671.

Marchesi JR, Adams DH, Fava F, Hermes GD, Hirschfield GM, Hold G *et al* (2015). The gut microbiota and host health: a new clinical frontier. *Gut*.

Mason OU, Scott NM (2014). Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill **8:** 1464-1475.

Medlin L, Simon N (1998). Phylogenetic Analysis of Marine Phytoplankton. In: Cooksey K (ed). *Molecular Approaches to the Study of the Ocean*. Springer Netherlands. pp 161-186.

Million M, Lagier JC, Yahav D, Paul M (2013). Gut bacterial microbiota and obesity. *Clin Microbiol Infect* **19:** 305-313.

Muhling M, Woolven-Allen J, Murrell JC, Joint I (2008). Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *Isme j* **2:** 379-392.

Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J (2014). Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS One* **9:** e94249.

Ohene-Adjei S, Teather RM, Ivan M, Forster RJ (2007). Postinoculation protozoan establishment and association patterns of methanogenic archaea in the ovine rumen. *Appl Environ Microbiol* **73:** 4609-4618.

Ojeda P, Bobe A, Dolan K, Leone V, Martinez K (2015). Nutritional modulation of gut microbiota - the impact on metabolic disease pathophysiology. *J Nutr Biochem*.

Olfert ED, Cross BM, McWilliams AA (1993). Guide to the care and use of experimental steers. *Canadian Council on Animal Care, Ottawa, ON, Canada*.

Petri RM, Schwaiger T, Penner GB, Beauchemin KA, Forster RJ, McKinnon JJ *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:** e83424.

Poulsen LK, Ballard G, Stahl DA (1993). Use of rRNA fluorescence in situ hybridization for measuring the activity of single cells in young and established biofilms. *Appl Environ Microbiol* **59:** 1354-1360.

Poulsen M, Schwab C, Jensen BB, Engberg RM, Spang A, Canibe N *et al* (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat Commun* **4:** 1428.

Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP *et al* (2007). The role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5:** 384-392.

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 Res* **41:** D590-D596.

R Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria.

Rius AG, Kittelmann S, Macdonald KA, Waghorn GC, Janssen PH, Sikkema E (2012). Nitrogen metabolism and rumen microbial enumeration in lactating cows with divergent residual feed intake fed high-digestibility pasture. *J Dairy Sci* **95:** 5024-5034.

Rooks MG, Veiga P, Wardwell-Scott LH, Tickle T, Segata N, Michaud M *et al* (2014). Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission **8:** 1403-1417.

Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM *et al* (2012). High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* **13:** 53.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537-7541.

Schwab C, Berry D, Rauch I, Rennisch I, Ramesmayer J, Hainzl E *et al* (2014). Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *Isme j* **8:** 1101-1114.

Schwieger F, Tebbe CC (1998). A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. *Appl Environ Microbiol* **64:** 4870-4876.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498-2504.

Sievert SM, Ziebis W, Kuever J, Sahm K (2000). Relative abundance of Archaea and Bacteria along a thermal gradient of a shallow-water hydrothermal vent quantified by rRNA slot-blot hybridization. *Microbiology* **146:** 1287-1293.

Snelling TJ, Genç B, McKain N, Watson M, Waters SM, Creevey CJ *et al* (2014). Diversity and community composition of methanogenic archaea in the rumen of scottish upland sheep assessed by different methods. *PLoS ONE* **9**.

Stark L, Giersch T, Wünschiers R (2014). Efficiency of RNA extraction from selected bacteria in the context of biogas production and metatranscriptomics. *Anaerobe* **29:** 85-90.

Stevenson DM, Weimer PJ (2007). Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl Microbiol Biotechnol* **75:** 165-174.

Tveit AT, Urich T, Svenning MM (2014). Metatranscriptomic analysis of arctic peat soil microbiota. *Appl Environ Microbiol* **80:** 5761-5772.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3:** e2527.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73:** 5261-5267.

Williams YJ, Popovski S, Rea SM, Skillman LC, Toovey AF, Northwood KS *et al* (2009). A vaccine against rumen methanogens can alter the composition of archaeal populations. *Appl Environ Microbiol* **75:** 1860-1866.

Wright ADG, Auckland CH, Lynn DH (2007). Molecular diversity of methanogens in feedlot cattle from Ontario and Prince Edward Island, Canada. *Appl Environ Microbiol* **73:** 4206-4210.

Xia Y, Kong Y, Seviour R, Yang HE, Forster R, Vasanthan T *et al* (2015). In situ identification and quantification of starch-hydrolyzing bacteria attached to barley and corn grain in the rumen of cows fed barley-based diets. *FEMS Microbiol Ecol* **91:** fiv077.

Yeoman CJ, White BA (2014). Gastrointestinal tract microbiota and probiotics in production animals. *Annu Rev Anim Biosci* **2:** 469-486.

2.6 Tables

Table 2.1 General sequence profiles of RNA-seq[1] and RNA/DNA Amplicon-seq[2] datasets

Method		Sample					Mean ± SEM
		L46	L52	L59	L132	L145	
RNA-seq	Bacterial 16S rRNA (%)	4.2	4.5	30.0	5.3	24.5	13.7±5.6
	Bacterial 23S rRNA (%)	7.7	9.3	50.6	9.4	42.5	23.9±9.3
	Archaeal 16S rRNA (%)	0.1	0.1	0.3	0.2	0.4	0.2±0.0
	Archaeal 23S rRNA (%)	0.4	0.2	0.8	0.9	1.5	0.7±0.2
	Eukaryotic 18S rRNA (%)	30.9	30.3	5.5	33.5	10.0	22.1±5.9
	Eukaryotic 28S rRNA (%)	47.8	46.3	9.2	42.3	15.5	32.2±8.2
	5S rRNA (%)	0.0	0.0	0.0	0.0	0.0	0.0±0.0
	5.8S rRNA (%)	0.1	0.1	0.0	0.1	0.0	0.1±0.0
	Other RNA (%)	8.8	9.2	3.7	8.4	5.6	7.1±1.1
	No. of total reads	42,161,316	41,860,772	41,422,056	33,763,140	33,273,904	38,496,238±2,037,011
RNA Amplicon-seq	No. of bacterial reads	7,331	8,513	8,472	6,991	5,798	7,421±506
	No. of archaeal reads	1,671	1,698	1,707	1,616	1,611	1,661±20
DNA Amplicon-seq	No. of bacterial reads	5,690	5,262	7,932	4,632	7,515	6,206±645
	No. of archaeal reads	1,157	1,177	1,684	1,286	1,359	1,333±95

1. RNA-seq: total RNA sequencing.

2. Amplicon-seq: sequencing of targeted PCR amplicons of bacterial and archaeal 16S rRNA/rDNA.

Table 2.2 Summary of sequences used for the taxonomic analysis from chimera-depleted RNA-seq and Amplicon-seq datasets

	RNA-seq			RNA Amplicon-seq			DNA Amplicon-seq		
	No. of reads	Classified	Unclassified	No. of reads	Classified	Unclassified	No. of reads	Classified	Unclassified
Bacteria (phylum)	2,413,786	94.6%	5.4%	30,644	98.2%	1.8%	25,526	98.8%	1.2%
Bacteria (family)		86.9%	13.1%		86.8%	13.2%		87.1%	12.9%
Archaeal (mixed)	25,312	98.0%	2.0%	7,741	98.6%	1.4%	6,041	99.7%	0.3%

Table 2.3 Differential taxa among RNA-seq and RNA/DNA Amplicon-seq datasets[1]

Taxa levels	Classifications	RNA-seq	RNA Amplicon-seq (Mean±SEM)	DNA Amplicon-seq	FDR[2]
Bacteria					
Phyla	*Bacteroidetes*	23.3±7.9%^{a}	22.7±8.1%^{a}	50.3±8.7%^{b}	0.09
Family	*Paraprevotellaceae*	0.2±0.1%^{a}	0.3±0.1%^{a}	1.8±0.3%^{b}	0.09
Family	*Prevotellaceae*	16.4±5.7%^{a}	17.9±7.2%^{a}	38.7±8.6%^{b}	0.14
Family	S24-7	0.2±0.1%^{a}	0.2±0.1%^{a}	3.1±1.0%^{b}	0.09
Phyla	*Firmicutes*[3]	16.2±4.5%	19.2±6.0%	29.4±6.2%	0.35
Family	*Lactobacillaceae*	0.3±0.1%^{a}	0.2±0.1%^{a}	8.1±5.0%^{b}	0.09
Family	*Mogibacteriaceae*	0.5±0.1%^{a}	0.8±0.2%^{ab}	1.3±0.2%^{b}	0.09
Phyla	*Lentisphaerae*	0.2±0.1%^{a}	1.3±0.6%^{b}	0.1±0.1%^{a}	0.09
Family	*Victivallaceae*	0.2±0.1%^{a}	1.3±0.6%^{b}	0.1±0.1%^{a}	0.09
Phyla	*Proteobacteria*	47.6±14.2%^{a}	46.3±14.3%^{a}	14.3±8.5%^{b}	0.09
Family	*Succinivibrionaceae*	45.6±14.0%^{a}	45.1±14.4%^{a}	13.8±8.6%^{b}	0.09
Phyla	*Synergistetes*	0.9±0.2%^{a}	3.6±1.1%^{b}	0.3±0.1%^{c}	0.12
Family	*Dethiosulfovibrionaceae*	0.8±0.2%^{a}	3.6±1.1%^{b}	0.3±0.1%^{c}	0.12
Archaea					
Mixed	*Methanomassiliicoccales*	38.8±11.9%^{a}	30.9±10.0%^{b}	15.7±2.0%^{ab}	0.09

1. Only commonly detected phylotypes among three datasets were compared.

2. *P* values were obtained using Repeated Measures ANOVA based on the arcsine square root-transformed proportion values, and then were adjusted into FDR using Benjamini-Hochberg method (Benjamini and Hochberg, 1995). A threshold of FDR<0.15 was applied to determine the significance. Within a row, means with different superscript are significantly different.

3. The relative abundance of the bacterial phylum *Firmicutes* did not show differences, but two families belonged to this phylum were different among datasets.

Table 2.4 A comparison of alpha-diversity estimators between RNA-seq and RNA/DNA Amplicon-seq datasets

	Bacteria			Archaea		
	RNA-seq	RNA Amplicon-seq (Mean±SEM)	DNA Amplicon-seq	RNA-seq	RNA Amplicon-seq (Mean±SEM)	DNA Amplicon-seq
Number of observed phylotypes	45.4±2.8$^{*1)}$	33.2±2.9$^{\#}$	35.6±3.1$^{\#}$	11.8±0.4^{a}	6.2±0.2^{b}	7.6±0.5^{c}
Chao	52.1±3.8*	43.0±7.3$^{*\#}$	40.4±3.3$^{\#}$	12.7±0.5^{a}	6.2±0.2^{b}	8.4±1.0^{b}
Shannon[2]	1.9±0.3	1.8±0.3	2.0±0.2	1.2±0.1	1.0±0.0	1.1±0.1
Inverse Simpson	4.3±1.2	4.2±1.3	4.8±1.0	2.5±0.2	2.3±0.1	2.5±0.3
Good's coverage	99.7%	99.8%	>99.8%	>99.8%	>99.9%	>99.9%

1. Within a row, means with different superscript tend to be different at P<0.1. Comparison was conducted using paired Wilcoxon signed rank test for bacterial and archaeal communities separately, and thus estimators between bacterial and archaeal groups are not comparable.

2. Shannon indices showed in the table are the raw value, and the comparison of Shannon index among datasets was based on the exponentially transformed values (Jost, 2007) using paired Wilcoxon signed rank test.

Table 2.5 qRT-PCR and qPCR primers

Target		Sequence (5'->3')	Reference
Total Bacteria	Forward	ACTCCTACGGGAGGCAG	(Stevenson and Weimer, 2007)
	Reverse	GACTACCAGGGTATCTAATCC	
Phylum *Bacteroidetes*	Forward	CAGCAGCCGCGGTAATAC	(Schwieger and Tebbe, 1998)
	Reverse	CCGTCAATTCCTTTGAGTTT	
Class *Gammaproteobacteria*	Forward	CMATGCCGCGTGTGTGAA	(Muhling *et al.*, 2008)
	Reverse	ACTCCCCAGGCGGTCDACTTA	

2.7 Figures

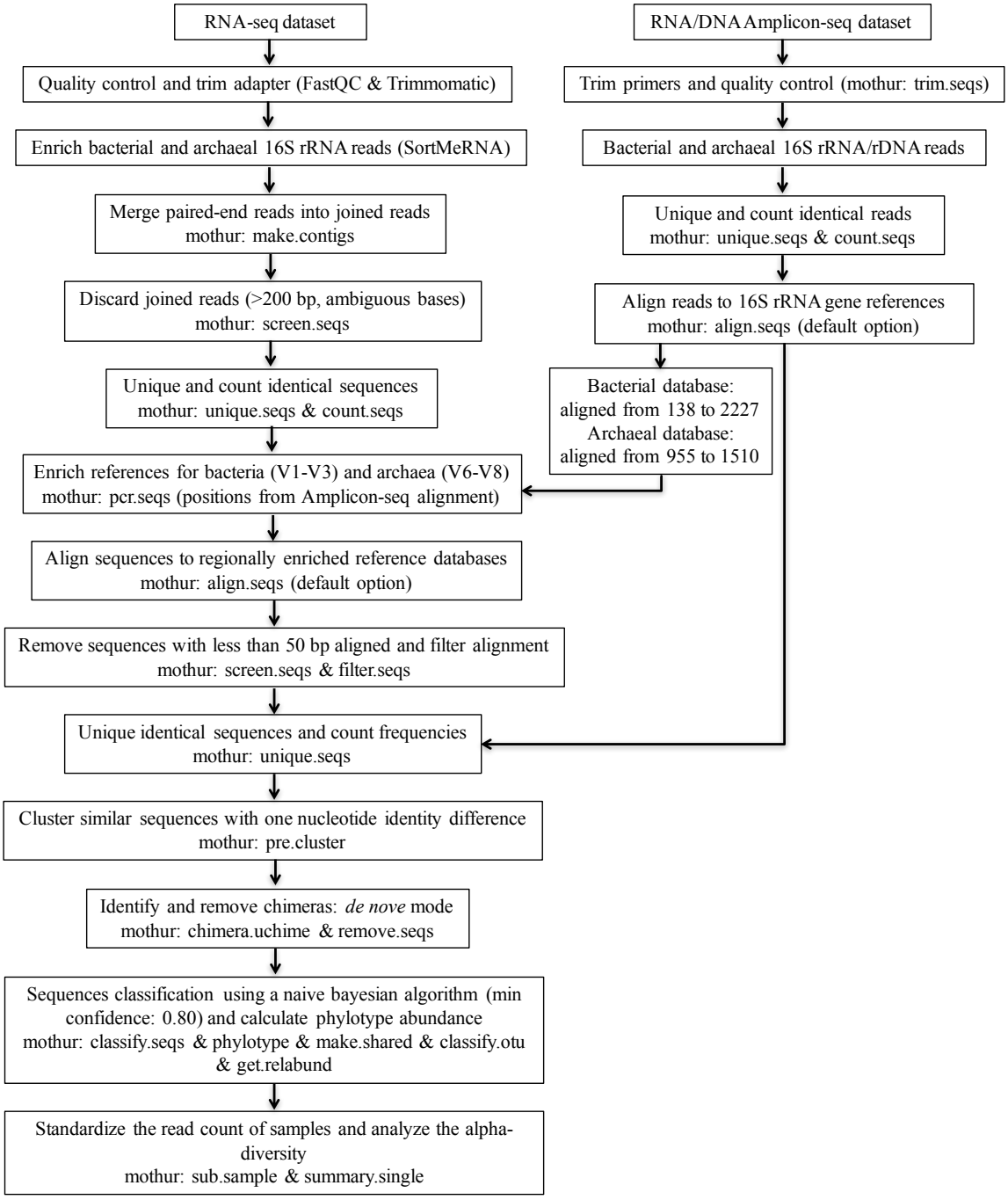


Figure 2.1 Flow chart of the pipeline for analyzing rumen microbiota using RNA-seq and RNA/DNA Amplicon-seq. The regionally enriched Greengenes 16S rRNA gene database (version gg_13_5_99 accessed May 2013) was used to analyze the bacterial community, and the regionally enriched rumen-specific archaeal database was used to analyze the archaeal community.

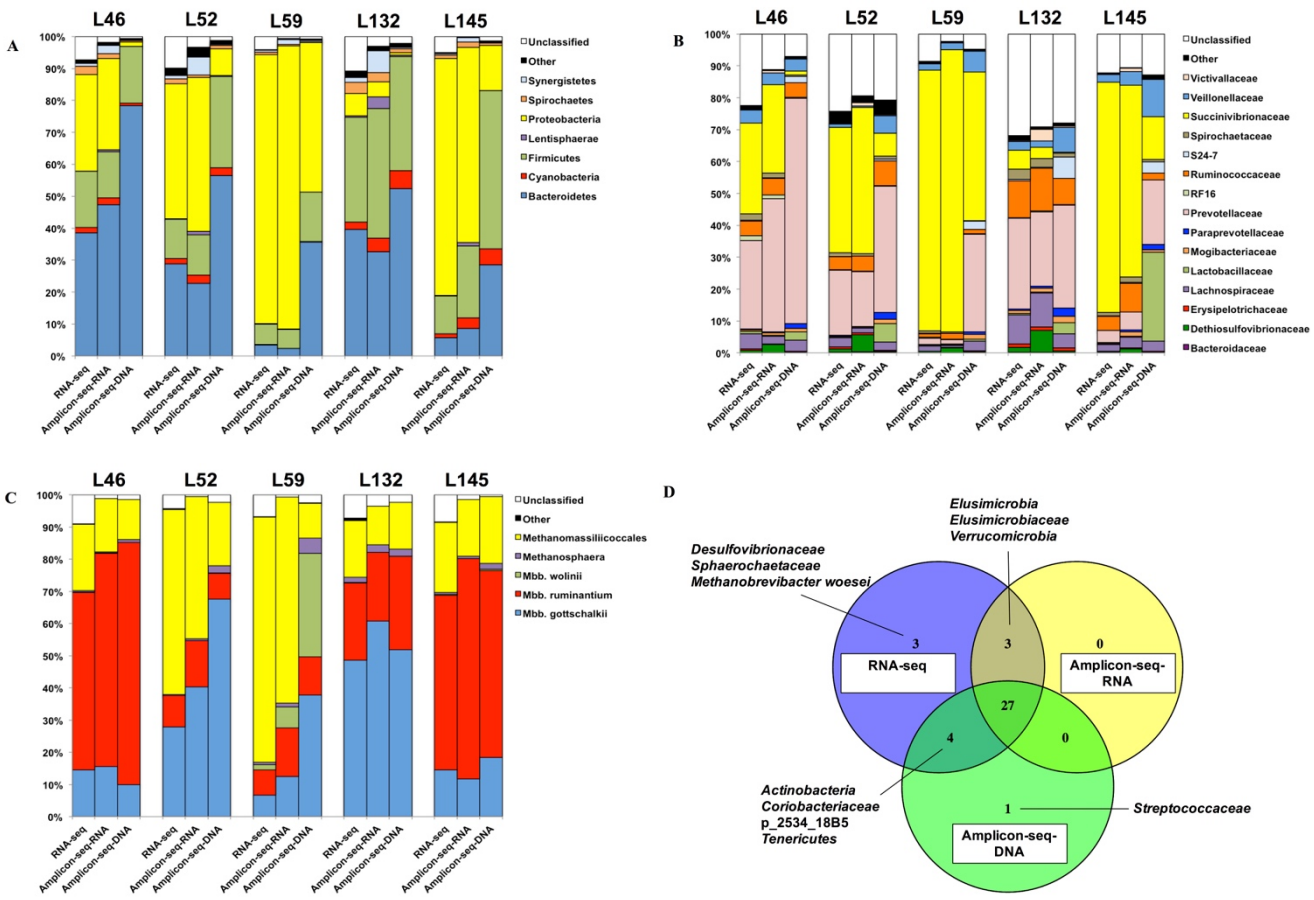


Figure 2.2 Microbial community composition estimated in the RNA-seq and RNA/DNA Amplicon-seq datasets. Microbial community composition of (**a**) bacterial phyla, (**b**) bacterial families, (**c**) archaea, and (**d**) dataset specific taxa.

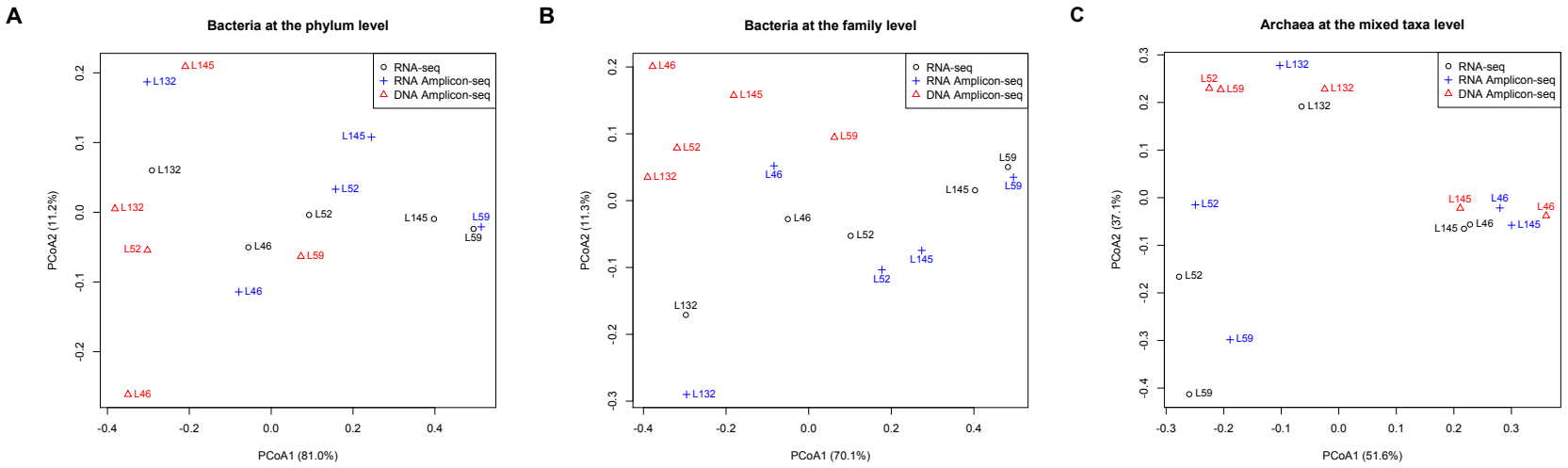


Figure 2.3 Dissimilarities among the RNA-seq, RNA Amplicon-seq and DNA Amplicon-seq datasets revealed by principal coordinate analysis (PCoA)**.** (**a**) PCoA based on shared bacterial phyla, (**b**) PCoA based on shared bacterial families, (**c**) PCoA based on shared archaeal mixed taxa. PCoA was performed based on the Bray-Curtis dissimilarity matrix.

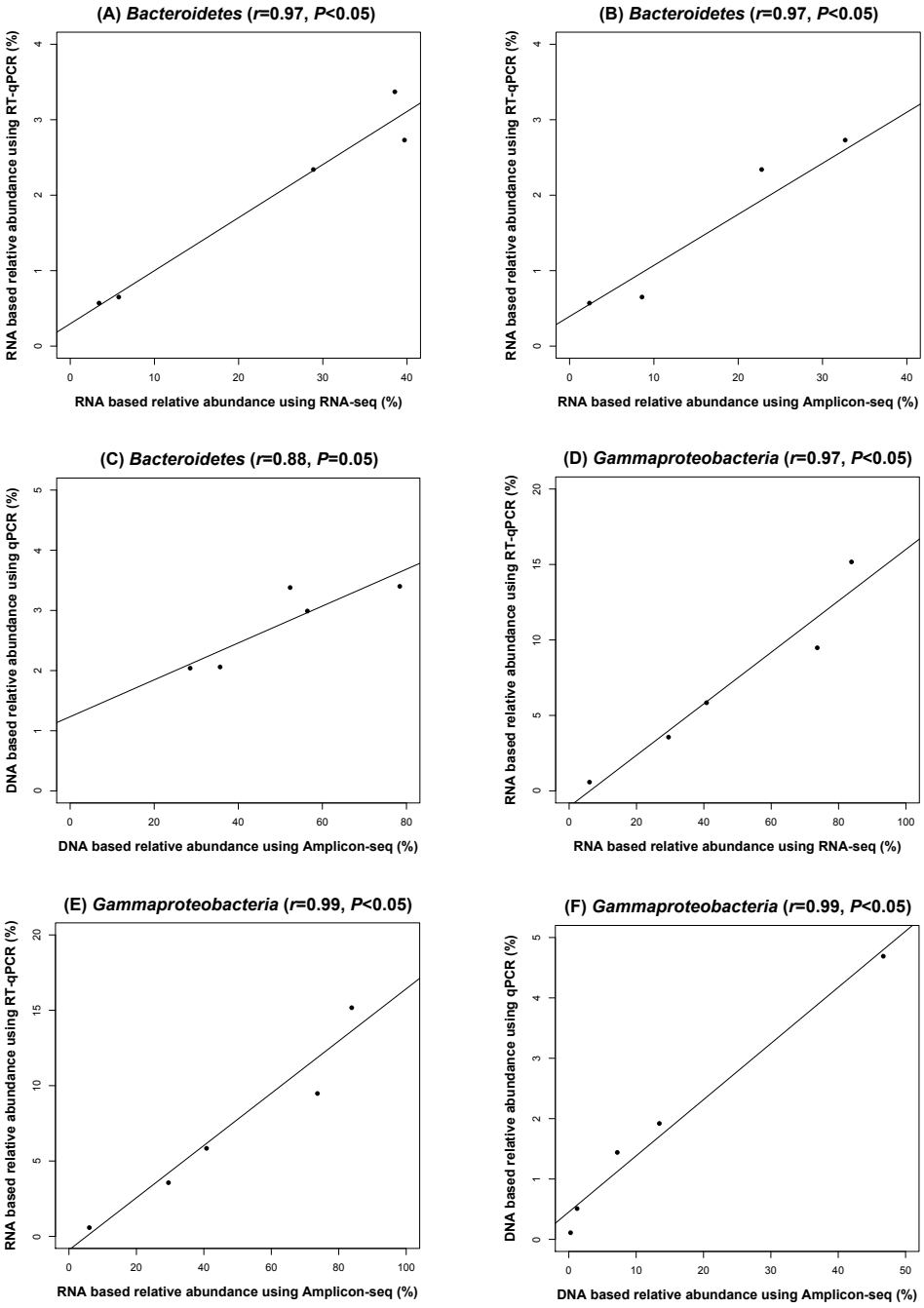


Figure 2.4 Validation of bacterial relative abundance using qRT-PCR and qPCR. Relative abundances of Bacteroidetes (**a**) and Gammaproteobacteria (**d**) from qRT-PCR and RNA-seq dataset, Bacteroidetes (**b**) and Gammaproteobacteria (**e**) from qRT-PCR and RNA Amplicon-seq dataset, and Bacteroidetes (**c**) and Gammaproteobacteria (**f**) from qPCR and DNA Amplicon-seq dataset.

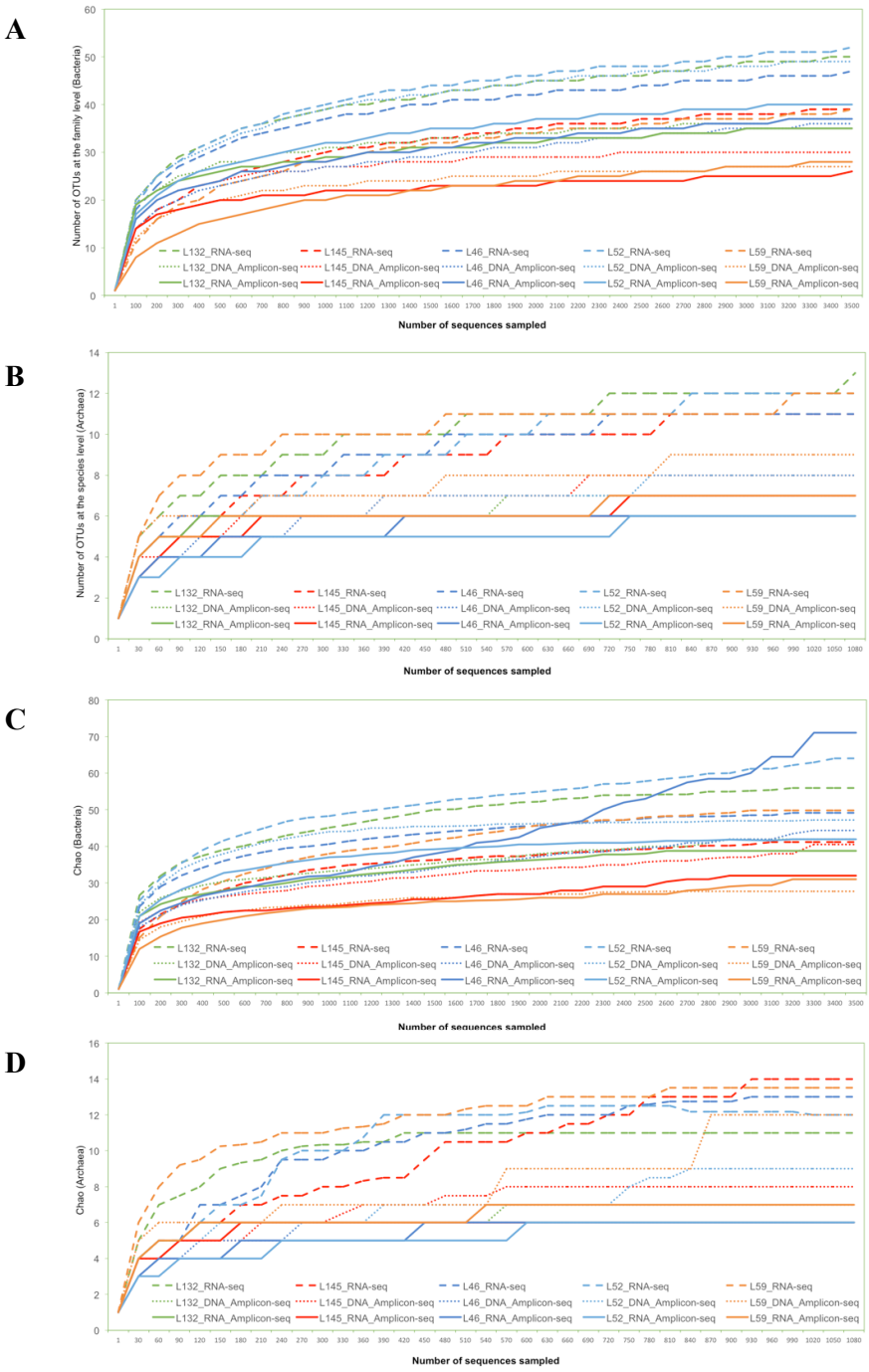


Figure 2.5 Rarefaction analysis of rumen bacteria and archaea. Collector's curves for (**a**) the number of bacterial phylotypes at the family level (**b**) the number of archaea phylotypes at the species level (**c**) Chao based on bacterial phylotypes at the family level (**d**) Chao based on archaeal phylotypes at the species level. The number of reads in each sample was randomly normalized to 3,476 for bacteria and 1,074 for archaea.

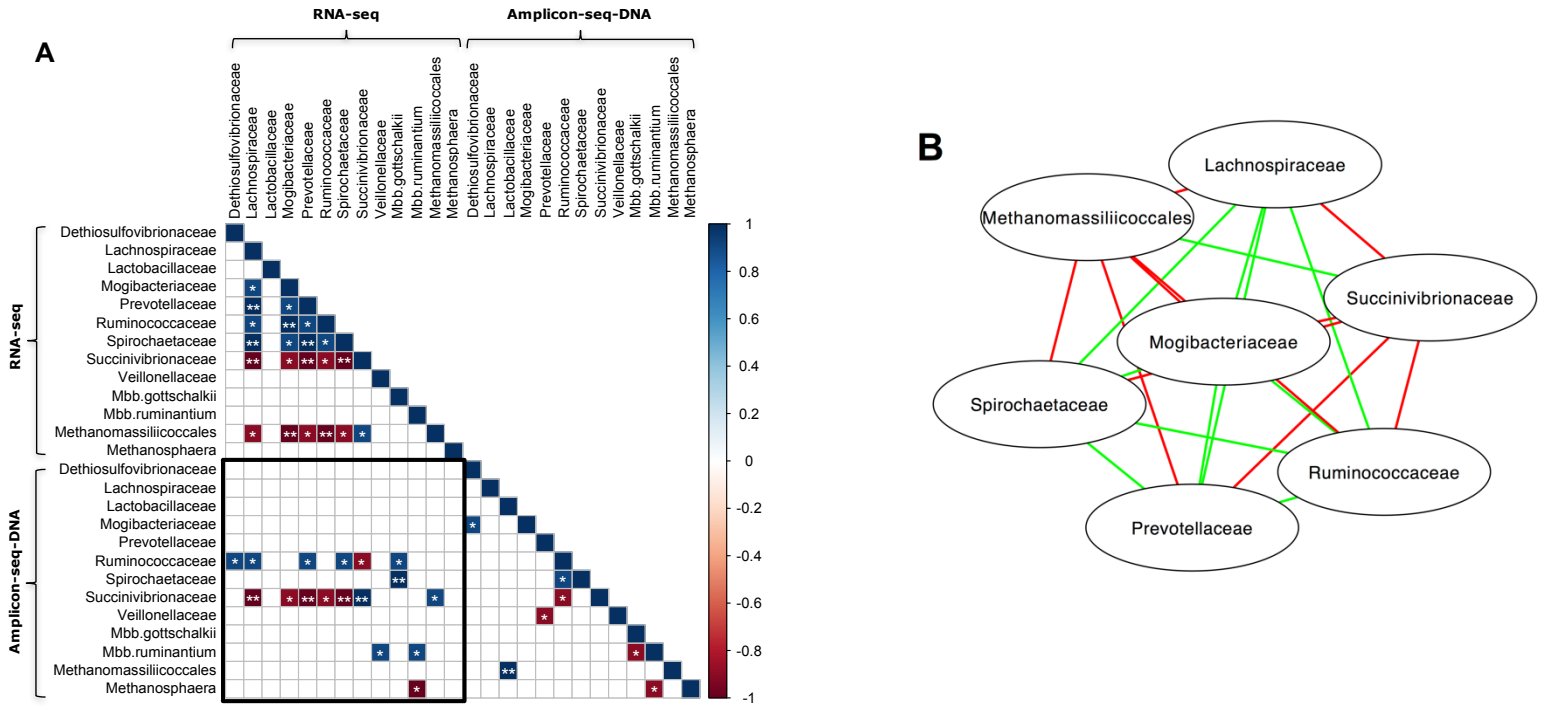


Figure 2.6 Co-occurrence of abundant microbial taxa in the RNA-seq and DNA Amplicon-seq datasets. (**a**) Correlation matrix of abundant microbial taxa and (**b**) Network of abundant microbial taxa in the RNA-seq dataset. Only bacterial families and archeael taxa with a relative abundance > 0.1% and detected in all five rumen samples using both RNA-seq and DNA Amplicon-seq were analyzed using the Spearman's rank correlation. The RNA Amplicon-seq dataset was not included into the analysis, because its bacterial and archaeal profiles were similar to profiles from the RNA-seq dataset. In (**a**), the sub-matrix surrounded by the black square exhibits correlations between taxa in the RNA-seq and DNA Amplicon-seq datasets. Strong correlations (Spearman's rank correlation coefficiency [r] ≥ 0.9 or ≤ -0.9) were displayed with * ($0.05 < P < 0.1$) and ** ($P < 0.05$), while the other correlations were showed as blank. In (**b**), a connection with a green/red line means a strong positive/negative correlation (r ≥ 0.9 or ≤ -0.9 and $P < 0.1$).

Chapter 3. Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle

3.1 Introduction

Ruminants can convert fibrous plant materials into edible meat and milk products for human consumption through rumen microbial fermentation. The symbiotic rumen microbiota mainly consists of bacteria, archaea, fungi, and ciliated protozoa (Deng *et al.*, 2008) that can produce volatile fatty acids (VFAs), microbial proteins, and vitamins for the host animals. Energy from VFAs (mainly acetate, propionate, and butyrate) produced by rumen microbes can fulfill approximately 70% of the host's energy requirements (Flint and Bayer, 2008), and microbial proteins can account for up to 90% of amino acids arriving into the small intestine of ruminants (Russell and Rychlik, 2001). In addition, rumen microbes can also synthesize most B-vitamins to meet the nutritional demands of the host (Santschi *et al.*, 2005). To date, multiple studies have reported that the rumen microbial composition is associated with feed efficiency in beef (Carberry *et al.*, 2012; Guan *et al.*, 2008; Hernandez-Sanabria *et al.*, 2010; Myer *et al.*, 2015a; Zhou *et al.*, 2009) and dairy (Jewell *et al.*, 2015) cattle, suggesting that rumen microbial function may be one of the factors that impact the cattle's feed efficiency. Feed efficiency is one of the most important traits in cattle farming because the feed cost accounts for most of the production expense, specifically, as 60 - 70% of the total cost in beef production (Karisa *et al.*, 2014). In addition, efficient beef cattle produce 20% less methane than inefficient ones (Nkrumah *et al.*, 2006); hence, enhancing the feed efficiency of beef cattle can also decrease their environmental footprint. Therefore, the improvement in feed efficiency is vital to meet the global food security needs of the next few decades.

Previous studies on associations between rumen microbiota and beef cattle feed efficiency revealed that the bacterial profiles in the rumen of efficient beef cattle differed from those of inefficient ones, and the abundances of bacterial genera such as *Butyrivibrio, Lactobacillus*, *Prevotella*, *Ruminococcus*, and *Succinivibrio* were associated with feed efficiency traits, including residual feed intake (RFI), dry matter intake (DMI), average daily gain (ADG), and feed conversion ratio (FCR) in beef steers (Guan *et al.*, 2008; Hernandez-Sanabria *et al.*, 2010; Myer *et al.*, 2015a) and heifers (Carberry *et al.*, 2012). Similarly, abundances of rumen archaeal taxa such as *Methanobrevibacter smithii*, *Methanosphaera stadtmanae*, and *Methanobrevibacter* sp. strain AbM4 have also been reported to be associated with RFI in beef steers and heifers (Carberry *et al.*, 2014; Zhou *et al.*, 2009). However, these studies only focused on the rumen microbial community composition and did not provide information on microbial metabolic functions.

Metagenomics and metatranscriptomics are approaches that can be used to study functional aspects of the microbial community at the genomic and transcriptional levels, respectively. Although metagenomics has been recently applied to link the rumen microbiome to host phenotypes (such as the feed efficiency and methane yields) in ruminants (Shabat *et al.*, 2016; Shi *et al.*, 2014; Wallace *et al.*, 2015), it does not allow for the assessment of the activities of rumen microbes or their actual gene expression levels. Until now, only a few studies have been performed to study the active functional characteristics of the rumen microbiome using metatranscriptomics, which includes the identification of eukaryotic gene expressions in muskoxen rumen (Qi *et al.*, 2011), the changes in rumen microorganisms and their gene expressions as related to sheep methanogenesis levels (Kamke *et al.*, 2016; Shi *et al.*, 2014), and

the activities and functions of the archaeal phylotype *Methanomassiliicoccales* in dairy cows in response to rapeseed oil (Poulsen *et al.*, 2013).

To date, linkages between active rumen microbiomes and feed efficiencies in beef cattle have not been studied yet. In the present study, we investigated the compositional and functional profiles of active rumen microbiomes in beef cattle with different RFIs using metatranscriptomics. RFI, the difference between an individual's actual feed intake and their predicted feed intake (Koch *et al.*, 1963; Nkrumah *et al.*, 2006), is one of the well-accepted feed efficiency measurements that has been applied in different livestock industries (Aggrey *et al.*, 2010; Koch *et al.*, 1963; Nkrumah *et al.*, 2006; Patience *et al.*, 2015). Animals with low RFI (L-RFI) are considered to be feed efficient, whereas high RFI (H-RFI) individuals are considered to be inefficient. Identification of rumen microbes that contribute to feed efficiency may provide information towards future applications for the direct administration of microbes to improve the feed efficiency of inefficient animals.

3.2 Materials and Methods

3.2.1 Animal experiments and rumen digesta sample collection

Animals used in this study were selected from a herd of 180 crossbred beef steers under feedlot conditions at the Kinsella Research Station at the University of Alberta. Animals received the high-energy finishing diet as described by Hernandez-Sanabria *et al.* (2013) from 10 months old till slaughter at 13 months old. The experimental protocol was approved by the Animal Care and Use Committee of the University of Alberta (protocol no. Moore-2006-55), according to the guidelines of the Canadian Council on Animal Care (Olfert *et al.*, 1993). RFI values were calculated based on DMI (dry matter intake), ADG (average daily gain), and MWT (metabolic

weight) during the 90-day experimental period, as described by Nkrumah *et al.* (2006). For each animal, rumen digesta was collected at slaughter, stored in RNA*later* (Ambion, Carlsbad, CA, USA) at 4 °C overnight, then transferred to -20 °C for long-term storage. In the current study, rumen digesta samples from 20 steers with the highest (n = 10) and lowest (n = 10) RFI values were subjected to RNA extraction and the metatranscriptomic analysis.

3.2.2 RNA extraction and sequencing

Total RNAs were extracted from rumen digesta samples using the procedure outlined in Chapter 2, which were modified based on the acid guanidinium-phenol-chloroform method (Béra-Maillet *et al.*, 2009; Chomczynski and Sacchi, 1987). For each sample, the RNA yield was determined using a Qubit 2.0 fluorimeter (Invitrogen, Carlsbad, CA, USA), and the RNA quality was examined with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA samples with RNA integrity number (RIN) scores higher than 7.0 were used for the metatranscriptomic analysis. Total RNA (100 ng) of each sample was used for the metatranscriptomic library construction using the TruSeq RNA sample Prep v2 LS kit (Illumina, San Diego, CA, USA), without the mRNA enrichment step. RNA libraries were evaluated using an Agilent 2200 TapeStation (Agilent Technologies) and a Qubit 2.0 fluorimeter (Invitrogen). cDNA fragments (~140 bp) were paired-end (2 × 100bp) sequenced using an Illumina HiSeq 2000 system at the McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada). All sequences were deposited into the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession number PRJNA275012.

3.2.3 Assessment of the active rumen microbiota using metatranscriptomics

Taxonomic assessment of the active rumen microbiota was conducted following the pipeline as described in Chapter 2. Briefly, after quality control and filtration of raw reads, bacterial and archaeal 16S rRNA sequences were selected and further aligned to the V1-V3 region-enriched Greengenes database (DeSantis *et al.*, 2006) and to the V6-V8 region-enriched rumen-specific archaea database (Janssen and Kirs, 2008), respectively, using mothur program (Schloss *et al.*, 2009). *De novo* chimera detection was performed, and chimera-removed sequences were taxonomically assessed using the naive Bayesian method. The microbial community diversity parameters were estimated, including the richness index (Chao), the evenness indices (Shannon Evenness and Simpson Evenness), and the alpha-diversity indices (Shannon and Inverse Simpson). To make these parameters comparable among metatranscriptomic libraries, the number of sequences per sample was normalized to the lowest counts for bacteria and archaea in all samples. Taxonomic profiles were summarized at the family level for bacteria and at the mixed taxonomic level for archaea.

3.2.4 Estimation of functional activities from rumen metatranscriptomes

To analyze functional pathways and carbohydrate-active enzymes (CAZymes) of the rumen metatransciptomes, putative mRNAs were obtained from metatransciptomic datasets after removing rRNAs using the SortMeRNA program (v1.9) (Kopylova *et al.*, 2012). These putative mRNAs were firstly aligned to the bovine genome (UMD 3.1) using TopHat 2.0.9 (Kim *et al.*, 2013) to filter out the host mRNAs. Then, the filtered mRNA sequences were *de novo* assembled for each sample using MetaVelvet (Namiki *et al.*, 2012) with multiple kmer sizes (k = 31, 41, 51, and 61) and a minimum output contig length of 100 bp. Assembled contigs from each sample

were then pooled together, and identical contigs were binned to construct a non-redundant contig set. Demultiplexed contigs were annotated using the UBLAST program in USEARCH v8.1.1861 (Edgar, 2010) against the KEGG database (Kanehisa *et al.*, 2012) for the metabolic pathway prediction, and against the Carbohydrate Active enZYme (CAZy) database (Lombard *et al.*, 2014) for the CAZymes analysis, separately. For each contig, the best hit with e-value < 1e-5, bitscore > 60, and sequence identity > 30% was used for annotation identification (Palomo *et al.*, 2016). For the metabolic pathway analysis, mRNAs were mapped to annotated contigs with KEGG information using the UBLAST program with an e-value of 1e-5 as the cut-off. Outputs of the mapping step were then used as inputs for the HUMAnN2 program (Abubucker *et al.*, 2012), to calculate the abundance of each pathway. To estimate CAZymes profiles, mRNAs were aligned to assembled contigs with CAZy annotation using bowtie2 program (Langmead and Salzberg, 2012) and quantified using SAMtools (Li *et al.*, 2009a). Abundances of pathways and CAZymes were normalized into counts per million reads (CPM) for downstream analysis.

3.2.5 Comparisons of rumen metatranscriptomes between H- and L-RFI beef steers

In this study, only bacterial families and archaeal taxa with a relative abundance greater than 0.1%, as well as metabolic pathways and CAZymes with CPM > 5, in at least 50% of the animals within each efficiency group were regarded as being detected and used for the comparison analysis. Relative abundance values of bacterial families and archaeal taxa were arcsine square root-transformed (Franzosa *et al.*, 2014) and were then subjected to comparative analysis between H- and L-RFI steers using the t-test. The data were transformed to stabilize the variance of the proportion data, making them more appropriate for application in the t-test framework. Principal coordinate analysis (PCoA) was also applied to virtualize the

compositional differences of the H- and L-RFI groups' rumen microbiota, which was conducted based on the Bray-Curtis dissimilarity matrices at the bacterial family level and at the archaeal species level. Moreover, the parsimony test and the analysis of molecular variance (AMOVA) between H- and L-RFI groups were performed to check whether their microbial communities had the same structure based on normalized metatranscriptomic datasets using mothur built-in code. Effective alpha-diversity parameters (Jost, 2007) were compared between these two groups using the Kruskal-Wallis rank sum test because these indices did not follow the normal distribution. To generally compare microbial functional profiles between efficient and inefficient animals, principal component analysis (PCA) was performed based on generalized logarithum transformed CPM of metabolic pathways and CAZymes separately. After that, the abundance of each microbial metabolic pathway and each CAZyme was compared between two animal groups using Linear discriminant analysis Effect Size (LEfSe) (Segata *et al.*, 2011), and features with Linear Discriminant Analysis (LDA) score > 2 and $P < 0.05$ were considered to be significantly different (Mottawea *et al.*, 2016).

Correlations between observed microbial taxa (bacterial families and archaeal taxa) and metabolic pathways were explored using Spearman's rank correlation. Correlation analyses were performed for all twenty animals, as well as within each RFI group. Correlations with $\rho > 0.5$ or < -0.5 and $P < 0.05$ were considered significant, with ρ representing the correlation coefficient.

3.3 Results

3.3.1 Rumen metatranscriptomes generated by RNA-seq

Residual feed intake (RFI) values were significantly different between H- and L-RFI groups (1.45 ± 0.17 vs. -1.64 ± 0.21 kg/day; P = 1.06e-9). In addition, H-RFI animals also had

higher DMI than L-RFI individuals (11.88 ± 0.33 vs. 8.79 ± 0.39 kg/day; *P* = 9.55e-6) (**Table 3.1**). A total of 1,202.5 million reads were generated from these 20 rumen samples with an average of 60,123,520 ± 1,870,616 (mean ± SEM) per library. After quality control, combining paired-end reads, clustering unique and similar sequences, and filtering chimeras, 673,171 ± 106,772 sequences were separated as bacterial V1-V3 regional 16S rRNAs, and 9,445 ± 1,380 sequences as archaeal V6-V8 regional 16S rRNAs. From each metatranscriptomic library, 4,595,730 ± 424,774 sequences were retained as mRNAs after removing host bovine mRNAs. After *de novo* assembly of mRNAs, it yielded an average of 302,815 ± 21851 contigs (N50 length: 155 ± 2 bp) and a total of 4,149,313 unique contigs. After aligning these contigs against KEGG and CAZy databases, 929,287 (22.4%) and 117,537 (2.8%) contigs were annotated, respectively, which were then used as the reference datasets. An average of 70.7 ± 2.2% and 3.2 ± 0.4% of mRNAs could be mapped back to these KEGG and CAZy annotated reference contigs, respectively. Details of the metatranscriptomic datasets are listed in **Table 3.2**.

3.3.2 Active rumen bacterial and archaeal communities of beef steers

Taxonomic profiling revealed that sequences belonging to ten bacterial phyla (ranging from 6 to 10), 19 bacterial families (ranging from 12 to 19), and six archaeal taxa (ranging from 4 to 6), accounted for 92.5 ± 0.8%, 80.0 ± 2.2%, and 91.7 ± 1.9% of the total bacterial and archaeal abundance, respectively (**Table 3.3**). Because 70.7% ± 3.8% of the bacterial reads could not be classified at the genus level, the downstream taxonomic analysis for the bacterial communities was retained at the family level. Based on Good's coverage values, the number of enriched 16S rRNA sequences from the metatranscriptomic datasets were sufficient to cover more than 99.0% of taxa at the bacterial family and the archaeal taxa levels (**Table 3.4**).

Bacterial and archaeal taxa with a relative abundance > 0.1% in all 20 animals were defined as the active core rumen microbiota. The active core microbiota consisted of six bacterial phyla including *Proteobacteria* (46.2 ± 6.1%), *Firmicutes* (23.2 ± 3.5%), *Bacteroidetes* (17.5 ± 2.7%), *Spirochaetes* (2.6 ± 0.4%), *Cyanobacteria* (1.0 ± 0.1%), and *Synergistetes* (0.7 ± 0.1%); eight bacterial families including *Succinivibrionaceae* (44.1 ± 6.2%), *Prevotellaceae* (11.2 ± 1.8%), *Ruminococcaceae* (7.6 ± 1.3%), *Lachnospiraceae* (6.1 ± 1.3%), *Veillonellaceae* (3.7 ± 0.9%), *Spirochaetaceae* (2.0 ± 0.3%), *Dethiosulfovibrionaceae* (0.7 ± 0.1%), and *Mogibacteriaceae* (0.7 ± 0.1%); and four archaeal taxa including *Methanomassiliicoccales* (35.6 ± 4.0%), *Methanobrevibacter ruminantium* (34.6 ± 4.4%), *Methanobrevibacter gottschalki* (18.4 ± 3.1%), and *Methanosphaera* (0.8 ± 0.1%). The core active microbiota accounted for 91.4 ± 0.9% (phylum level), 76.0 ± 2.6% (family level), and 89.4 ± 3.3% (taxa level) of the bacterial and archaeal abundance, respectively. Although these core taxa were consistently detected among the 20 animals, we observed noticeable individual variations in their abundances (coefficient of variation [CV], ranged from 50.2% to 104.0%) (**Table 3.3**).

3.3.3 Active microbial metabolic functions and enzymes in the rumen of beef cattle

Based on metabolic pathway profiles generated using HUMAnN2, 189 KEGG pathways were observed, and 57 of them were excluded from downstream analyses as exogenous pathways. In total, 112 core pathways were ubiquitous in all 20 rumen samples with CPM > 5, which represented 72.9 ± 1.0% of mapped mRNAs (**Figure 3.1a**). These 112 core pathways belonged to four first level KEGG functional categories, including “metabolism” (33.4 ± 1.2%), “genetic information processing” (31.7 ± 1.0%), “cellular processes” (6.6 ± 0.3%), and “environmental information processing” (1.2 ± 0.2%). At the second level of the KEGG hierarchy, these core

pathways belonged to 20 KEGG functional categories, with "carbohydrate metabolism" (14.1 ± 0.4%), "translation" (11.4 ± 0.8%), "folding, sorting and degradation" (10.9 ± 0.4%), "replication and repair" (9.1 ± 0.7%), and "amino acid metabolism" (5.2 ± 0.3%) as the top five most abundant functions. Within these five second level functional groups, "ko00030: Pentose phosphate pathway" (2.5 ± 0.1%), "ko03010: Ribosome" (10.0 ± 0.7%), "ko04120: Ubiquitin mediated proteolysis" (6.2 ± 0.5%), "ko03430: Mismatch repair" (3.3 ± 0.3%), and "ko00250: Alanine, aspartate and glutamate metabolism" (2.3 ± 0.1%) were the most abundant pathways in their respective functional groups. In addition, eukaryotic pathways were also identified in our datasets, such as "ko04145: phagosome" (1.6 ± 0.1%), "ko04142: lysosome" (1.4 ± 0.1%), and so on (**Figure 3.1a**). In terms of CAZyme profiles, a total of 168 CAZymes were detected (6 AA [auxiliary activity], 14 CE [carbohydrate esterase], 91 GH [glycoside hydrolase], 46 GT [glycosyl transferase], and 11 PL [polysaccharide lyase]), and 126 of them were observed with CPM > 5 in all 20 animals, which were considered as core rumen CAZymes (representing 94.8 ± 4.7% of total CAZyme reads). GT2 (33.0 ± 5.7%), GH13 (5.8 ± 0.6%), and GH9 (5.8 ± 0.7%) were the most abundant CAZymes in the rumen of these steers (**Figure 3.1b**).

3.3.4 Comparison of active rumen microbiota between H- and L-RFI beef steers

The principal coordinate analysis (PCoA) did not show clear separations of the active rumen microbiota between H- and L-RFI steers based on the Bray-Curtis dissimilarity matrices of bacterial families and archaeal taxa (**Figure 3.2**). Meanwhile, the parsimony test and the analysis of molecular variance (AMOVA) also did not show significant differences between the H- and L-RFI groups in either the bacterial and archaeal communities (P = 0.34 and 0.44 for the bacterial community, respectively; P = 0.67 and 0.13 for the archaeal community). When

diversity indices were compared, the richness index (Chao), the evenness indices (Shannon Evenness and Simpson Evenness), and the alpha-diversity indices (Shannon and Inverse Simpson) were not significantly different ($P > 0.05$, Kruskal-Wallis rank sum test) (**Table 3.4**) between the two groups. However, when the relative abundance of each taxon was compared, three bacterial families *Lachnospiraceae*, *Lactobacillaceae*, *Veillonellaceae* and one archaeal taxon *Methanomassiliicoccale* tended to be different (P = 0.08, 0.08, 0.09 and 0.07, t-test, respectively) between the H- and L-RFI cattle (**Table 3.2**). The relative abundance of *Lachnospiraceae, Lactobacillaceae*, and *Veillonellaceae* in the H-RFI group (8.3 ± 2.3%, 0.6 ± 0.1%, and 5.2 ± 1.6%, respectively) was at least 2-fold higher than those in the L-RFI group (3.8 ± 0.7%, 0.3 ± 0.1%, and 2.2 ± 0.4%, respectively). The relative abundance of *Methanomassiliicoccales* was higher in L-RFI steers (42.9 ± 6.6%) than that in H-RFI steers (28.3 ± 3.5%) (**Table 3.2**).

3.3.5 Differential microbial metabolic pathways and CAZymes between H- and L-RFI cattle

Although the PCA did not show clear separation based on CAZyme profiles between two feed efficiency groups, it showed two major clusters according to RFI classification based on detected KEGG pathways, except for four individuals (L37, L52, L59, and H50) (**Figure 3.3a** and **3.3b**). Through conducting the LEfSe analysis, 32 differentially expressed KEGG pathways were identified (LDA score > 2 and $P < 0.05$) between the two animal groups: 30 pathways were more abundant in the H-RFI group, while two pathways (ko03015: mRNA surveillance pathway and ko04130: SNARE interactions in vesicular transport) were more abundant in L-RFI animals (**Figure 3.3c** and **3.3d**). Among these 32 differentially abundant pathways, 26, 4, and 2 of them

were related to “metabolism”, “genetic information processing”, and “cellular processes” (first level KEGG functions), respectively. Within these 26 metabolism-related pathways, there were six “cofactors and vitamin metabolism” pathways (ko00670, ko00730, ko00770, ko00785, ko00790, and ko00860), five “terpenoids and polyketides metabolism” pathways (ko00281, ko00900, ko00908, ko01051, and ko01055), and four “amino acid metabolism” pathways (ko00270, ko00300, ko00340, and ko00350; related to cysteine, histidine, lysine, methionine, and tyrosine). Meanwhile, the LEFSe analysis also revealed 12 differentially abundant CAZymes between two RFI groups (LDA score > 2 and $P < 0.05$). Only one CAZyme (GT31) was enriched in L-RFI animals, and the other 11 CAZymes (AA0, CE4, CE14, GH14, GH36, GT4, GT8-9, GT28, GT30, and GT47) were all higher abundant in the inefficient group (**Figure 3.3e** and **3.3f**).

3.3.6 Relationships between active phylotypes and metabolic pathways in the rumen

Correlation analysis using Spearman's rank correlation revealed various significant (coefficient [ρ] > 0.5 or < -0.5 and $P < 0.05$) relationships between active taxa (bacterial families and archaeal taxa) and active metabolic pathways (**Figure 3.4a)**. When animals from two feed efficiency groups were analyzed together, in total of 115 significant correlations were identified. However, when relationships between taxa and metabolic pathways were further explored within each RFI group, more correlations were revealed (n = 177 in the H-RFI group, and n = 152 in the L-RFI group) and correlation patterns were different between two RFI groups (**Figure 3.4b** and **3.4c**). Moreover, the relative abundances of three differential bacterial families between H- and L-RFI animals (*Lachnospiraceae*, *Lactobacillaceae*, *Veillonellaceae*), were all significantly correlated with the same four pathways (ko00680: Methane metabolism, ko00630: Glyoxylate

and dicarboxylate metabolism, ko00380: Tryptophan metabolism, and ko00280: Valine leucine and isoleucine degradation) and clustered together in the H-RFI group, while their relative abundances were associated with different pathways in L-RFI animals.

3.4 Discussion

The current study aimed to link the active rumen microbiome with feed efficiency (RFI) in beef cattle by assessment of the active rumen microbiota and its functions using total RNA-based metatranscriptomics. Firstly, active core rumen microbiota including six bacterial phyla, eight bacterial families, and four archaeal taxa were identified, suggesting their essential roles in occupying rumen ecological niches because they have also been reported to inhabit the rumen of beef heifers (Petri *et al.*, 2013), lactating cows (Jami and Mizrahi, 2012), and other ruminant species (Henderson *et al.*, 2015) using DNA-based identification. For example, family *Prevotellaceae* (belonging to phylum *Bacteroidetes*) is highly abundant at both the DNA (up to 59.9%) (Stevenson and Weimer, 2007) and RNA levels (11.2%, this study). Members of *Prevotellaceae* utilize various substrates, such as starch, protein, peptides, hemicellulose, and pectin, to generate a wide range of end products, mainly short chain fatty acids including acetate, succinate, and propionate (Carberry *et al.*, 2012; Russell and Rychlik, 2001). This bacterial family has a remarkable degree of genetic diversity (Avgustin *et al.*, 1997; Purushe *et al.*, 2010), which occupies various ecological niches within the rumen (Jami and Mizrahi, 2012). Our results provide further understanding regarding its consistent presence in the rumen, especially with animals fed a barley-based high grain diet. It was surprising to identify the high relative abundance (7.6%) of the family *Ruminococcaceae* (belonging to phylum *Firmicutes*) in the active core rumen microbiota under a barley-based high grain diet because this family has been

considered to consist of mainly fibrolytic organisms. However, recent studies have reported that members of this family such as *Ruminococcus bromii* and other *Ruminococcus* spp. are highly involved in starch hydrolysis (Klieve *et al.*, 2007; Klieve *et al.*, 2012), and the relative abundance of this family in the rumen is associated with the proportion of barley grain in the diet (Xia *et al.*, 2015). Due to the considerable genetic diversity among members in this family (Klieve *et al.*, 2007), it is necessary to look at them at genus, species, and/or strain level to further understand the functions of this family in the rumen under high grain dietary conditions.

Different from previous studies, we found that the most abundant active rumen bacterial family in beef steers fed with barley was *Succinivibrionaceae* (44.1 ± 6.2%), which belongs to the phylum *Proteobacteria*. Previous DNA-based studies revealed its abundance ranging from 0.6 to 20.1% (Kim and Yu, 2014; Petri *et al.*, 2013). Members of *Succinivibrionaceae* can utilize hydrogen to produce succinate, a precursor of propionate (Pope *et al.*, 2011), one of the major VFAs utilized for gluconeogenesis in the liver (Young, 1977). The predominance of *Succinivibrionaceae* at the RNA level suggests it may be the major contributor for rumen propionate production under the barley-based high grain diet through production of upstream succinate. Future studies to compare the abundance of this family under different dietary conditions (different types of grain, high grain vs. high forage), as well as to measure the concentration of succinate, will provide more direct information of its functional role in the rumen. In the meantime, we identified two archaeal taxa, *Methanomassiliicoccales* and *Mbb. ruminantium*, as the most active methanogens in the rumen of steers fed a barley diet. *Methanomassiliicoccales* utilize methanol and methylamines as major energy and carbon sources to produce methane (Poulsen *et al.*, 2013; Sollinger *et al.*, 2016), while *Mbb. ruminantium* produces CH_4 by utilizing CO_2, formate and H_2 as substrates (Miller *et al.*, 1986; Russell and

Rychlik, 2001). Their high abundances suggest that the methyl compound conversion route and the CO_2 reduction route are the major methanogenesis pathways in the rumen of beef steers fed a barley-based diet.

The further identification of 112 active core metabolic pathways and 126 CAZymes in this study provides more understanding of the basic housekeeping activities of the rumen microbiome. The high abundance of metabolism-related functions (carbohydrate, amino acid, and energy metabolism), especially carbohydrate metabolism, confirmed similar findings reported using metagenomic analysis (Wang *et al.*, 2013). For example, "ko00500: starch and sucrose metabolism" was observed as one of the core pathways in the rumen of beef cattle fed with grain diet based on metagenomic datasets (Wang *et al.*, 2013). In addition to starch and sucrose metabolism, this pathway is also related to the metabolism of many hydrolytic products (e.g., maltose, glucose, fructose, xylose, etc.), which are by-products of the digestion of feed starch (Wang *et al.*, 2013); therefore, it is not surprising to detect the high activity of this pathway in beef cattle fed with barley. Meanwhile, "ko00680: methane metabolism" was also ubiquitous in all 20 animals, which further highlights the importance of methane production as part of the key microbial functions in rumen. Previous studies reported that although genes belonging to this pathway had similar abundance between high and low methane emitters (Shi *et al.*, 2014; Wallace *et al.*, 2015), variations in their expression levels were associated with methane yields (Shi *et al.*, 2014). The consistent presence and the noticeable abundance variations (CV: 21.9%) of this pathway at the RNA level further suggest that it could be the target for the rumen methane mitigation as suggested by Shi *et al.* (2014). Our findings underline some of the housekeeping functions of rumen microbiomes, including microbial metabolism and

survival, and their variations may contribute to different host phenotypic performances such as cattle feed efficiency.

Indeed, differences in rumen metatranscriptomes at both taxonomic and functional levels between H- (inefficient) and L-RFI (efficient) cattle were identified. Different relative abundance of *Lachnospiraceae* between RFI groups suggests that the activity of this family may be associated with feed efficiency in beef cattle, probably because some members of *Lachnospiraceae* are major butyrate producers (Meehan and Beiko, 2014). Recently, ruminal butyrate concentration has been reported to be associated with feed efficiency (Guan *et al.*, 2008), and the abundance of this family at the DNA level has also been linked to feed efficiency (Myer *et al.*, 2015a) in beef steers. Therefore, the higher abundance of this family in H-RFI animals may be accompanied by increased butyrate metabolism, which affects feed efficiency. However, the identified more butyrate producers in inefficient animals is in conflict with the previous findings that L-RFI animals tended to have higher butyrate concentration as compared to H-RFI animals (Guan *et al.*, 2008). Such discrepancy may be due to that the abundances being identified at the family level, while this family has 24 genera and several unclassified strains with different functional niches (Sayers *et al.*, 2011), and less than half of its members have the butyrate production capacity (Meehan and Beiko, 2014). Moreover, we did not detect differences in butyrate concentration in the rumen fluid of the same twenty animals (Hernandez-Sanabria *et al.*, 2010) and found no significant correlation between the activity of *Lachnospiraceae* and the "ko00650: butanoate metabolism" pathway. Further studies at deeper taxonomic levels (genus/species) and/or using culture-based methods may provide more information regarding the relationships between *Lachnospiraceae*, butyrate production, and feed efficiency.

Additionally, *Methanomassiliicoccales* was more abundant in L-RFI animals. The members belonging to this order are H_2-dependent methylotrophic methanogens, which can use methylamine as energy and carbon sources and provide additional NH_4^+ to the entire rumen bacteria (Brugere *et al.*, 2014; Poulsen *et al.*, 2013). The concentration of methylamine in the rumen increases with increasing amounts of dietary barely, and methylamine is a harmful component for the host due to its degradation products such as hydrogen peroxide and formaldehyde (Ametaj *et al.*, 2010; Saleem *et al.*, 2012; Yu *et al.*, 2006). The higher abundance of *Methanomassiliicoccales* in efficient animals suggests that more methylamine might be utilized to supply additional NH_4^+ during the methanogenesis which may not only generate NH_4^+ for rumen nitrogen cycling but also decrease the potential negative effects caused by high concentrations of methylamine.

In the current study, we found that 30 differential metabolic pathways and 11 CAZymes were more abundant in H-RFI cattle, while only 2 pathways and 1 CAZymes were more abundant in the L-RFI group. A recent metagenomics study also revealed that a significantly higher number of KEGG pathways were enriched in the microbiomes of H-RFI cattle (Shabat *et al.*, 2016), suggest that the rumen microbiomes of inefficient cattle may have higher and more diverse activities than those of efficient cattle. In addition, four amino acid metabolism pathways related to cysteine, histidine, lysine, methionine, and tyrosine were more active in H-RFI beef steers. Using metagenomics, Shabat *et al.* (2016) also reported that these four and several other pathways related to protein/amino-acid metabolism were enriched in H-RFI dairy cows. These data suggest differences in nitrogen metabolism between animals with different feed efficiencies. Specifically, rumen microbiomes in H-RFI animals seem to have higher nitrogen metabolism activities. In line with our speculation, previous studies have shown that H-RFI dairy cows and

beef heifers had low rumen NH_3 concentration (Lawrence *et al.*, 2011; Rius *et al.*, 2012), indicating higher rumen nitrogen utilization (Bach *et al.*, 2005; Broderick and Reynal, 2009) as compared to L-RFI individuals. Furthermore, different correlation patterns between microbial phylotypes and functional pathways were observed for two RFI groups, meaning that the same phylotypes may have different functional niches in the rumen of H- and L-RFI animals. One taxon may involve multiple metabolic pathways and one metabolic pathway may be attributable to multiple taxa. These complicated relationships could not be revealed by our correlation analysis. Furthermore, as reviewed and discussed by Blazewicz *et al.* (2013), there are limitations of using rRNA as the proxy of microbial abundances. Future culture-based studies need to be applied to explore relationships among rRNA abundances, mRNA abundances, microbial taxa abundances, and metabolic functions to validate our current findings.

Although the PCA based on metabolic pathways generated two major clusters corresponding to steers' RFI rankings, there are four animals as exceptions. This indicates a role for the host in the relationship between rumen microbiome and feed efficiency. It has been reported that host genetics and sire breed can impact the association between rumen microbiota and RFI (Guan *et al.*, 2008; Hernandez-Sanabria *et al.*, 2013). Future studies to identify the host factors regulating the rumen microbiome are needed to further elucidate how host-microbial interactions in the rumen contribute to the beef cattle feed efficiency.

Together, the above metatranscriptome analysis indicate that inefficient (H-RFI) cattle's microbiomes may have higher activities and more diverse abilities to utilize feed compounds and to generate more diverse products. The higher microbial activity in H-RFI animals could be due to their higher intake. It is known that microbial activities are strongly correlated with the feed intake (Pathak, 2008) because rumen microorganisms can have more available substrates and

nutrients for growth. In the meantime, increased feed intake can lead to a higher rumen passage rate (Colucci *et al.*, 1982), which is associated with increased microbial growth and activities (Martinez *et al.*, 2009; Shriver *et al.*, 1986), due to a decrease in the proportion of the energy used for maintenance and reduced microbial generation times (Firkins, 1996; Sniffen and Robinson, 1987). In addition, the difference in activity outcomes could be also affected by the total microbial mass, which was not evaluated in this study due to the limitation in sampling process. Moreover, efficiency of rumen function is related to both microbial production and host absorption. Although the rumen microbiomes of H-RFI animals may be more active, the host may have less absorption capability and metabolism. Indeed, transcriptome profiles of ruminal epithelial tissues of the same animals revealed higher expression of genes involved in intercellular gaps in L-RFI steers, indicating greater and more efficient paracellular absorption of nutrients by efficient animals as compared to inefficient animals (Kong *et al.*, 2016). Therefore, future studies on measuring metabolites and/or enzymes in the rumen, total rumen microbial mass as well as rumen passage rate are needed to explore and further understand the relationships among microbial fermentation, host rumen epithelial absorption, and feed efficiency.

3.5 Conclusions

This study comprehensively explored the active compositional and functional characteristics of the rumen microbiome in beef cattle using metatranscriptomics. Our results revealed an active core microbiome in the bovine rumen, which showed consistent occurrences but noticeable variations among individuals. The presence of a rumen microbial core indicates that there are fundamental and essential active components commonly involved in rumen

functions. Furthermore, cattle with different RFIs showed different rumen microbial characteristics at both the compositional and functional levels, suggesting possible links between the active rumen microbiome and host feed efficiency. It is noticeable that only a small fraction of assembled mRNA contigs (22.4% in the datasets) could be mapped to the KEGG database, and the functional analysis only focused on these annotated mRNAs in the present study. Similarly, Ross *et al.* (2012) reported that only 6.0% of metagenomic reads could be mapped to existing databases. A low ratio of mapped reads highlights that the majority of rumen microbial genes have not been deposited in current databases and that bovine rumen microbial functions have yet to be sufficiently addressed. Potentially, the Hungate 1000 project (http://www.rmgnetwork.org/hungate1000.html) will provide a reference dataset of rumen microbial genome sequences after sequencing cultivated rumen bacteria, archaea, fungi and ciliate protozoa. To date, the Hungate 1000 project has 437 rumen microbial cultures being sequenced and information of 311 genomes are already available in the Joint Genome Institute website (JGI; http://genome.jgi.doe.gov). Future studies through comprehensive annotation using culture-based references will help identify the unmapped mRNAs. Regardless, our findings provide evidence that the active rumen microbiome is one of the biological factors that may contribute to the variations in feed efficiency of beef cattle. The information from the current study highlights the possibility to enhance nutrient utilization and improve feed efficiency through altering rumen microbial functions.

3.6 References

Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL *et al* (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* **8:** e1002358.

Aggrey SE, Karnuah AB, Sebastian B, Anthony NB (2010). Genetic properties of feed efficiency parameters in meat-type chickens. *Genet Sel Evol* **42:** 25.

Ametaj BN, Zebeli Q, Saleem F, Psychogios N, Lewis MJ, Dunn SM *et al* (2010). Metabolomics reveals unhealthy alterations in rumen metabolism with increased proportion of cereal grain in the diet of dairy cows. *Metabolomics* **6:** 583-594.

Avgustin G, Wallace RJ, Flint HJ (1997). Phenotypic diversity among ruminal isolates of Prevotella ruminicola: proposal of Prevotella brevis sp. nov., Prevotella bryantii sp. nov., and Prevotella albensis sp. nov. and redefinition of Prevotella ruminicola. *Int J Syst Bacteriol* **47:** 284-288.

Bach A, Calsamiglia S, Stern MD (2005). Nitrogen metabolism in the rumen. *J Dairy Sci* **88 Suppl 1:** E9-21.

Béra-Maillet C, Mosoni P, Kwasiborski A, Suau F, Ribot Y, 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. *J Microbiol Methods* **77:** 8-16.

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *Isme j* **7:** 2061-2068.

Broderick GA, Reynal SM (2009). Effect of source of rumen-degraded protein on production and ruminal metabolism in lactating dairy cows. *J Dairy Sci* **92:** 2822-2834.

Brugere JF, Borrel G, Gaci N, Tottey W, O'Toole PW, Malpuech-Brugere C (2014). Archaebiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut Microbes* **5:** 5-10.

Carberry CA, Kenny DA, Han S, McCabe MS, Waters SM (2012). Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. *Appl Environ Microbiol* **78:** 4949-4958.

Carberry CA, Waters SM, Kenny DA, Creevey CJ (2014). Rumen methanogenic genotypes differ in abundance according to host residual feed intake phenotype and diet type. *Appl Environ Microbiol* **80:** 586-594.

Chomczynski P, Sacchi N (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162:** 156-159.

Colucci PE, Chase LE, Van Soest PJ (1982). Feed Intake, Apparent Diet Digestibility, and Rate of Particulate Passage in Dairy Cattle. *J Dairy Sc* **65:** 1445-1456.

Deng W, Xi D, Mao H, Wanapat M (2008). The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. *Mol Biol Rep* **35:** 265-274.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069-5072.

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-2461.

Firkins JL (1996). Maximizing microbial protein synthesis in the rumen. *J Nutr* **126:** 1347s-1354s.

Flint HJ, Bayer EA (2008). Plant cell wall breakdown by anaerobic microorganisms from the Mammalian digestive tract. *Ann N Y Acad Sci* **1125:** 280-288.

Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM *et al* (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* **111:** E2329-2338.

Guan LL, Nkrumah JD, Basarab JA, Moore SS (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol Lett* **288:** 85-91.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH (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.

Hernandez-Sanabria E, Guan LL, Goonewardene LA, Li M, Mujibi DF, Stothard P *et al* (2010). Correlation of particular bacterial PCR-denaturing gradient gel electrophoresis patterns with bovine ruminal fermentation parameters and feed efficiency traits. *Appl Environ Microbiol* **76:** 6338-6350.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Zhou M, Moore SS, Guan LL (2013). Influence of sire breed on the interplay among rumen microbial populations inhabiting the rumen liquid of the progeny in beef cattle. *PLoS One* **8:** e58461.

Jami E, Mizrahi I (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS One* **7:** e33306.

Janssen PH, Kirs M (2008). Structure of the archaeal community of the rumen. *Appl Environ Microbiol* **74:** 3619-3625.

Jewell KA, McCormick CA, Odt CL, Weimer PJ, Suen G (2015). Ruminal Bacterial Community Composition in Dairy Cows Is Dynamic over the Course of Two Lactations and Correlates with Feed Efficiency. *Appl Environ Microbiol* **81:** 4697-4710.

Jost L (2007). Partitioning diversity into independent alpha and beta components. *Ecology* **88:** 2427-2439.

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:** 56.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40:** D109-114.

Karisa B, Moore S, Plastow G (2014). Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle. *Anim Sci J* **85:** 374-387.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14:** R36.

Kim M, Yu Z (2014). Variations in 16S rRNA-based microbiome profiling between pyrosequencing runs and between pyrosequencing facilities. *J Microbiol* **52:** 355-365.

Klieve AV, O'Leary MN, McMillen L, Ouwerkerk D (2007). Ruminococcus bromii, identification and isolation as a dominant community member in the rumen of cattle fed a barley diet. *J Appl Microbiol* **103:** 2065-2073.

Klieve AV, McLennan SR, Ouwerkerk D (2012). Persistence of orally administered Megasphaera elsdenii and Ruminococcus bromii in the rumen of beef cattle fed a high grain (barley) diet. *Anim Prod Sci* **52:** 297-304.

Koch RM, Swiger LA, Chambers D, Gregory KE (1963). Efficiency of Feed Use in Beef Cattle. *J Anim Sci* **22:** 486-494.

Kong RS, Liang G, Chen Y, Stothard P, Guan le L (2016). Transcriptome profiling of the rumen epithelium of beef cattle differing in residual feed intake. *BMC Genomics* **17:** 592.

Kopylova E, Noe L, Touzet H (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28:** 3211-3217.

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357-359.

Lawrence P, Kenny DA, Earley B, Crews DH, Jr., McGee M (2011). Grass silage intake, rumen and blood variables, ultrasonic and body measurements, feeding behavior, and activity in pregnant beef heifers differing in phenotypic residual feed intake. *J Anim Sci* **89:** 3248-3261.

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078-2079.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42:** D490-495.

Martinez ME, Ranilla MJ, Ramos S, Tejido ML, Carro MD (2009). Effects of dilution rate and retention time of concentrate on efficiency of microbial growth, methane production, and ruminal fermentation in Rusitec fermenters. *J Dairy Sci* **92:** 3930-3938.

Meehan CJ, Beiko RG (2014). A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol Evol* **6:** 703-713.

Miller TL, Wolin MJ, Zhao HX, Bryant MP (1986). Characteristics of methanogens isolated from bovine rumen. *Appl Environ Microbiol* **51:** 201-202.

Mottawea W, Chiang CK, Muhlbauer M, Starr AE, Butcher J, Abujamel T *et al* (2016). Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. *Nat Commun* **7:** 13419.

Myer PR, Smith TP, Wells JE, Kuehn LA, Freetly HC (2015). Rumen microbiome from steers differing in feed efficiency. *PLoS One* **10:** e0129174.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* **40:** e155.

Nkrumah JD, Okine EK, Mathison GW, Schmid K, Li C, Basarab JA *et al* (2006). Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle. *J Anim Sci* **84:** 145-153.

Olfert ED, Cross BM, McWilliams AA (1993). Guide to the care and use of experimental steers. *Canadian Council on Animal Care, Ottawa, ON, Canada*.

Palomo A, Jane Fowler S, Gulay A, Rasmussen S, Sicheritz-Ponten T, Smets BF (2016). Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp. *Isme j* **10:** 2569-2581.

Pathak AK (2008). Various factors affecting microbial protein synthesis in the rumen. *Vet World* **1:** 186-189.

Patience JF, Rossoni-Serao MC, Gutierrez NA (2015). A review of feed efficiency in swine: biology and application. *J Anim Sci Biotechnol* **6:** 33.

Petri RM, Schwaiger T, Penner GB, Beauchemin KA, Forster RJ, McKinnon JJ *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:** e83424.

Pope PB, Smith W, Denman SE, Tringe SG, Barry K, Hugenholtz P *et al* (2011). Isolation of Succinivibrionaceae implicated in low methane emissions from Tammar wallabies. *Science* **333:** 646-648.

Poulsen M, Schwab C, Jensen BB, Engberg RM, Spang A, Canibe N *et al* (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat Commun* **4:** 1428.

Purushe J, Fouts DE, Morrison M, White BA, Mackie RI, Coutinho PM *et al* (2010). Comparative genome analysis of Prevotella ruminicola and Prevotella bryantii: insights into their environmental niche. *Microb Ecol* **60:** 721-729.

Qi M, Wang P, O'Toole N, Barboza PS, Ungerfeld E, Leigh MB *et al* (2011). Snapshot of the eukaryotic gene expression in muskoxen rumen-A metatranscriptomic approach. *PLoS ONE* **6**.

Rius AG, Kittelmann S, Macdonald KA, Waghorn GC, Janssen PH, Sikkema E (2012). Nitrogen metabolism and rumen microbial enumeration in lactating cows with divergent residual feed intake fed high-digestibility pasture. *J Dairy Sci* **95:** 5024-5034.

Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM *et al* (2012). High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* **13:** 53.

Russell JB, Rychlik JL (2001). Factors that alter rumen microbial ecology. *Science* **292:** 1119-1122.

Saleem F, Ametaj BN, Bouatra S, Mandal R, Zebeli Q, Dunn SM *et al* (2012). A metabolomics approach to uncover the effects of grain diets on rumen health in dairy cows. *J Dairy Sci* **95:** 6606-6623.

Santschi DE, Berthiaume R, Matte JJ, Mustafa AF, Girard CL (2005). Fate of supplementary B-vitamins in the gastrointestinal tract of dairy cows. *J Dairy Sci* **88:** 2043-2054.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K *et al* (2011). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **39:** D38-51.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537-7541.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS *et al* (2011). Metagenomic biomarker discovery and explanation. *Genome Biol* **12:** R60.

Shabat SK, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Berg Miller ME *et al* (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *Isme j*.

Shi W, Moon CD, Leahy SC, Kang D, Froula J, Kittelmann S *et al* (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res* **24:** 1517-1525.

Shriver BJ, Hoover WH, Sargent JP, Crawford RJ, Jr., Thayne WV (1986). Fermentation of a High Concentrate Diet as Affected by Ruminal pH and Digesta Flow. *J Dairy Sci* **69:** 413-419.

Sniffen CJ, Robinson PH (1987). Protein and fiber digestion, passage, and utilization in lactating cows. Microbial growth and flow as influenced by dietary manipulations. *J Dairy Sci* **70:** 425-441.

Sollinger A, Schwab C, Weinmaier T, Loy A, Tveit AT, Schleper C *et al* (2016). Phylogenetic and genomic analysis of Methanomassiliicoccales in wetlands and animal intestinal tracts reveals clade-specific habitat preferences. *FEMS Microbiol Ecol* **92**.

Stevenson DM, Weimer PJ (2007). Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl Microbiol Biotechnol* **75:** 165-174.

Wallace RJ, Rooke JA, McKain N, Duthie CA, Hyslop JJ, Ross DW *et al* (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* **16:** 839.

Wang L, Hatem A, Catalyurek UV, Morrison M, 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:** e78507.

Xia Y, Kong Y, Seviour R, Yang HE, Forster R, Vasanthan T *et al* (2015). In situ identification and quantification of starch-hydrolyzing bacteria attached to barley and corn grain in the rumen of cows fed barley-based diets. *FEMS Microbiol Ecol* **91:** fiv077.

Young JW (1977). Gluconeogenesis in cattle: significance and methodology. *J Dairy Sci* **60:** 1-15.

Yu PH, Lu LX, Fan H, Kazachkov M, Jiang ZJ, Jalkanen S *et al* (2006). Involvement of semicarbazide-sensitive amine oxidase-mediated deamination in lipopolysaccharide-induced pulmonary inflammation. *Am J Pathol* **168:** 718-726.

Zhou M, Hernandez-Sanabria E, Guan LL (2009). Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Appl Environ Microbiol* **75:** 6524-6533.

3.7 Tables

Table 3.1 Residual feed intake, dry matter intake, and RNA quality of samples

ID	*RFI*	*DMI*	*RIN*	*ID*	*RFI*	*DMI*	*RIN*
	High RFI (H-RFI) group				*Low RFI (L-RFI) group*		
H11	1.74	13.48	8.7	L37	-1.39	9.39	9.6
H16	2.75	13.14	9.2	L40	-1.57	9.51	8.2
H17	0.98	11.49	10	L46	-1.10	9.91	8.7
H27	1.23	11.81	8.0	L52	-2.03	10.12	7.7
H50	1.11	12.01	8.8	L59	-2.13	8.70	9.4
H56	1.63	13.14	8.9	L96	-1.22	8.89	8.2
H107	1.64	11.05	8.6	L132	-2.33	7.08	7.2
H150	1.43	11.06	7.9	L145	-2.81	6.24	8.8
H152	0.85	10.51	9.1	L166	-1.11	8.79	7.8
H176	1.11	11.12	9.2	L167	-0.75	9.25	9.2
Mean	1.45	11.88			-1.64	8.79	
SEM	0.17	0.33			0.21	0.39	
RFI			P = 1.06e-9 (t-test)				
DMI			P = 9.55e-6 (t-test)				

Abbreviations: RFI, residual feed intake; DMI, dry matter intake; RIN, RNA integrity number.

Table 3.2 Details of metatranscriptomic datasets

ID	*RNA reads before QC*[1] *(n)*	*RNA reads after QC*[1] *(n)*	*Bacterial 16S rRNAs (n)*	*Archaeal 16S rRNAs (n)*	*Bacterial 16S V1-3 rRNAs*[2] *(n)*	*Archaeal 16S V6-8 rRNAs*[2] *(n)*	*Putative mRNAs (n)*	*Filtered mRNAs*[3] *(n)*	*Contigs from assembly (n)*	*N50 length (bp)*	*mRNAs mapped to KEGG contigs (%)*	*mRNAs mapped to CAZy contigs (%)*
H11	73827508	57354074	889564	56032	151643	8773	8163684	8161850	372803	162	80.1	1.8
H16	79194124	60906642	1992501	76529	355507	11689	8581748	8579902	374521	160	80.2	2.8
H17	65401858	47977240	7447570	196346	1213447	26179	3325950	3324960	461132	182	44.8	7.1
H27	62235658	45957158	7629938	46312	1517769	6836	2486024	2485536	225135	149	62.0	5.3
H50	56771000	41494434	6847953	69968	1262589	9660	2504632	2501746	246297	159	58.1	6.1
H56	69151956	51379420	5269009	86009	974614	13755	5357570	5353220	475048	153	69.3	2.7
H107	61481768	46533042	2283981	98971	385906	14846	5557232	5485382	350655	152	75.2	1.8
H150	58736994	45041246	2530302	99000	455280	14912	5329762	5326836	421865	154	70.1	1.9
H152	59751554	44557518	5838899	75394	1077754	10269	3017982	3017154	204919	153	68.7	3.8
H176	56635428	42618418	4258673	135999	700214	18598	3920272	3918724	353586	156	65.1	3.5
L37	64079478	45866208	7616850	52640	1450595	7413	2327054	2325998	206300	152	61.8	5.8
L40	60218214	46117206	1170583	16835	213773	2326	6101582	6097016	376242	152	75.9	1.7
L46	54205486	41288720	879337	19785	160778	2853	5446990	5445430	193744	157	85.4	1.8
L52	53474730	40956170	920779	11455	166953	1588	5548290	5545406	329771	152	76.3	1.5
L59	55264142	40562424	6073409	50022	1149642	6952	2128574	2127674	142198	150	68.6	5.5
L96	62808446	47472578	2677871	70125	460620	9814	4950778	4939332	343411	150	71.6	1.9
L132	43604698	32822248	873288	34116	136487	4279	4140402	4138846	206500	151	82.5	1.5
L145	44825366	32706490	4005649	64819	737083	9029	2462640	2370336	207088	151	64.3	3.5
L166	61866444	46318910	639976	5561	114850	824	6617426	6614742	331424	155	77.9	1.9
L167	58935556	44182504	3980271	58490	777920	8299	4156582	4154512	233664	152	76.2	2.8
Mean	60123520	45105633	3691320	66220	673171	9445	4606259	4595730	302815	155	70.7	3.2
SEM	1870616	1483393	567906	9983	106772	1380	424073	424774	21851	2	2.2	0.4

1. QC: quality control.
2. These numbers are after chimera removing.
3. These numbers are after removing bovine mRNAs.

Table 3.3 Relative abundances of active bacterial and archaeal taxa estimated using metatranscriptomics

Taxonomic level	*Classification*	*Overall*	*H-RFI* (Mean±SEM)	*L-RFI*	*P*[1]
Bacteria[1]					
Phyla	*Actinobacteria*	0.2±0.0%	0.3±0.0%	0.2±0.0%	0.14
Family	*Coriobacteriaceae*	0.2±0.0%	0.2±0.0%	0.2±0.0%	0.18
Phyla	*Bacteroidetes*	17.6±2.7%	16.6±3.3%	18.6±4.4%	0.80
Family	*Bacteroidaceae*	0.1±0.0%	0.2±0.1%	0.1±0.0%	0.75
Family	*p-2534-18B5*	0.9±0.3%	1.2±0.6%	0.6±0.3%	0.64
Family	*Paraprevotellaceae*	0.2±0.0%	0.2±0.1%	0.3±0.1%	0.72
Family	*Prevotellaceae*	11.2±1.8%	9.9±1.9%	12.4±3.2%	0.64
Family	*RF16*	0.3±0.1%	0.2±0.1%	0.5±0.2%	0.16
Family	*S24-7*	0.2±0.0%	0.2±0.0%	0.2±0.0%	1.00
Phyla	*Cyanobacteria*	1.0±0.1%	1.0±0.2%	1.1±0.2%	0.83
Phyla	*Firmicutes*	23.2±3.5%	28.0±6.2%	18.5±3.0%	0.21
Family	*Erysipelotrichaceae*	0.8±0.2%	0.7±0.2%	0.9±0.3%	0.63
Family	*Lachnospiraceae*	6.1±1.3%	8.3±2.3%	3.8±0.7%	0.08
Family	*Lactobacillaceae*	0.5±0.1%	0.6±0.1%	0.3±0.1%	0.08
Family	*Mogibacteriaceae*	0.7±0.1%	0.7±0.2%	0.6±0.1%	0.66
Family	*Ruminococcaceae*	7.6±1.3%	8.8±2.0%	6.4±1.6%	0.33
Family	*Veillonellaceae*	3.7±0.9%	5.2±1.6%	2.2±0.4%	0.09
Phyla	*Lentisphaerae*	0.2±0.0%	0.2±0.1%	0.2±0.1%	0.91
Family	*Victivallaceae*	0.2±0.0%	0.2±0.1%	0.2±0.0%	0.94
Phyla	*Proteobacteria*	46.2±6.1%	42.6±9.4%	49.8±8.3%	0.56
Family	*Desulfovibrionaceae*	0.2±0.0%	0.3±0.1%	0.2±0.0%	0.56
Family	*Succinivibrionaceae*	44.1±6.2%	40.8±9.4%	47.5±8.3%	0.57
Phyla	*Spirochaetes*	2.6±0.4%	3.0±0.6%	2.2±0.5%	0.41
Family	*Sphaerochaetaceae*	0.3±0.1%	0.5±0.2%	0.2±0.0%	0.11
Family	*Spirochaetaceae*	2.0±0.3%	2.2±0.5%	1.9±0.5%	0.71
Phyla	*Synergistetes*	0.7±0.1%	0.7±0.1%	0.7±0.1%	0.95
Family	*Dethiosulfovibrionaceae*	0.7±0.1%	0.7±0.1%	0.7±0.1%	0.95
Phyla	*Tenericutes*	0.2±0.0%	0.2±0.0%	0.2±0.1%	0.91
Phyla	*Verrucomicrobia*	0.5±0.1%	0.4±0.2%	0.6±0.2%	0.43
Archaea[1]					
Mixed	*Methanobrevibacter gottschalki*	18.4±3.1%	17.2±4.7%	19.7±4.1%	0.56
Mixed	*Methanobrevibacter ruminantium*	34.6±4.4%	41.6±6.0%	27.5±5.8%	0.11
Mixed	*Methanobrevibacter woesei*	0.1±0.0%	0.1±0.0%	0.1±0.0%	0.83
Mixed	*Methanobrevibacter wolinii*	2.2±1.4%	2.7±2.5%	1.7±1.3%	0.96
Mixed	*Methanomassiliicoccales*	35.6±4.0%	28.3±3.5%	42.9±6.6%	0.07
Mixed	*Methanosphaera*	0.8±0.1%	0.7±0.1%	0.8±0.1%	0.88

1. *P*-value between the H- and L-RFI groups was obtained using the t-test based on the arcsine square root-transformed relative abundance values.
2. Bacterial and archaeal communities were analyzed separately.

Table 3.4 Comparisons of alpha-diversity indices[1] between H- and L-RFI cattle

	Bacteria				Archaea			
	Overall (n=20)	H-RFI (n=10) (Mean±SEM)	L-RFI (n=10)	P^2	Overall (n=20)	H-RFI (n=10) (Mean±SEM)	L-RFI (n=10)	P^2
Richness index								
Chao1	96.75±2.17	95.65±2.72	97.86±3.48	0.65	10.35±0.22	10.19±0.34	10.51±0.28	0.36
Evenness indices								
Shannon evenness	0.45±0.03	0.46±0.05	0.43±0.04	0.41	0.55±0.02	0.56±0.02	0.54±0.03	1.00
Simpson Evenness	0.06±0.01	0.07±0.01	0.05±0.01	0.23	0.30±0.01	0.31±0.01	0.29±0.02	0.94
Diversity indices								
Shannon	1.98±0.15	2.04±0.22	1.92±0.20	0.50	1.22±0.04	1.23±0.05	1.21±0.07	0.94
Inverse Simpson	5.16±0.71	5.89±1.18	4.44±0.79	0.29	2.78±0.14	2.84±0.19	2.72±0.21	0.94
Good's coverage[3]	>99%	>99%	>99%		>99%	>99%	>99%	

1. To make alpha-diversity indices comparable among metatranscriptomic libraries, the number of sequences per library was normalized to 114,850 and 824 for bacteria community and archaea community, respectively, according to the minimum read number among all libraries.
2. *P*-value was obtained based on effective values of each index (transformed according to Jost (2007)) using the Kruskal-Wallis rank sum test.
3. Good's coverage = $(1-n_1/N) \times 100\%$, where the n_1 is the number of OTUs that only sampled once and N is the number of all OTUs.

3.8 Figures

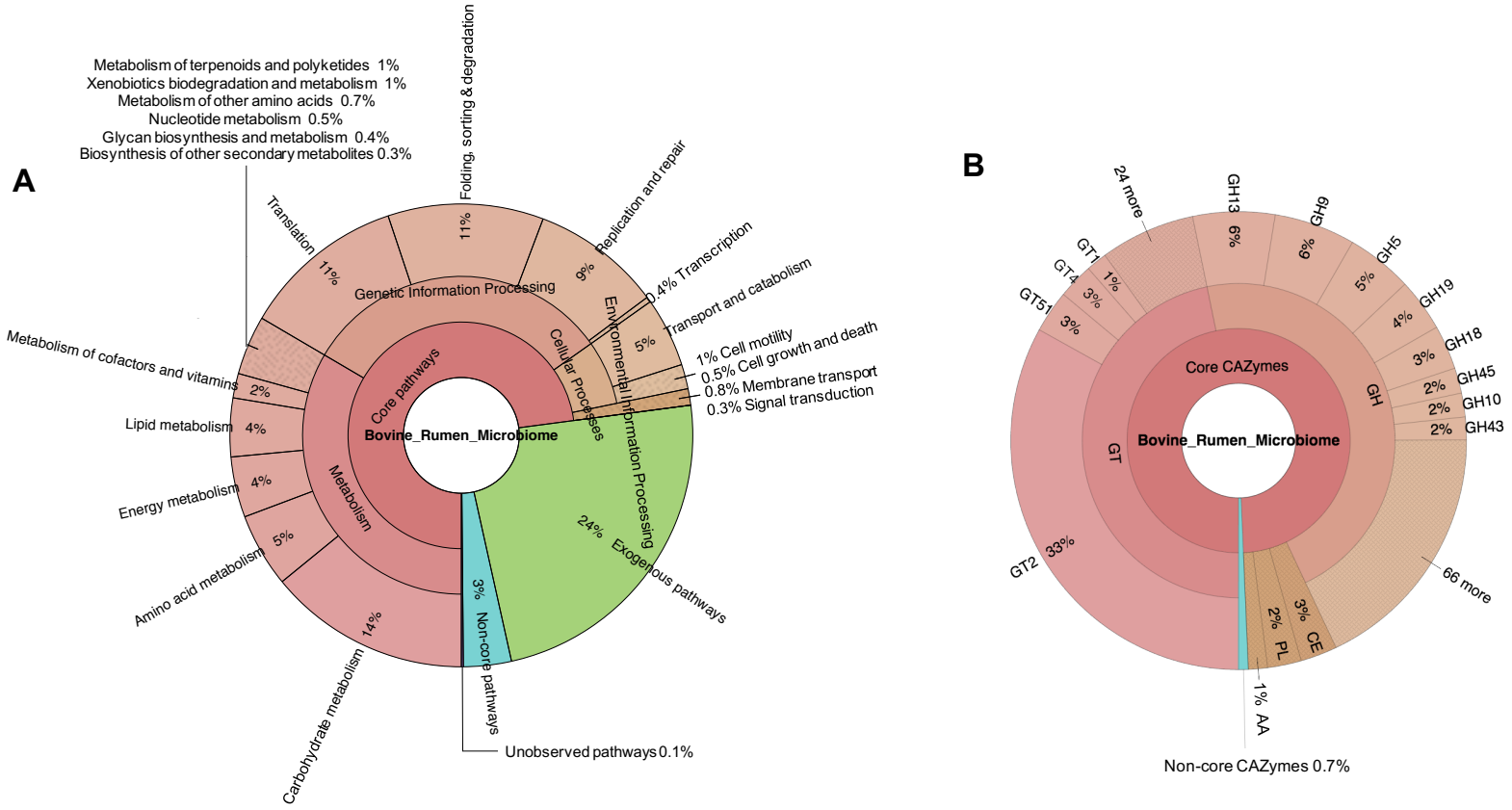


Figure 3.1 Profiles of bovine rumen microbiome. (**a**) Microbial metabolic pathways based on their First and Second Level functions in the KEGG hierarchy. (**b**) Carbohydrate-active enzymes (CAZymes). These graphs were generated using the program Krona (Ondov *et al.*, 2011).

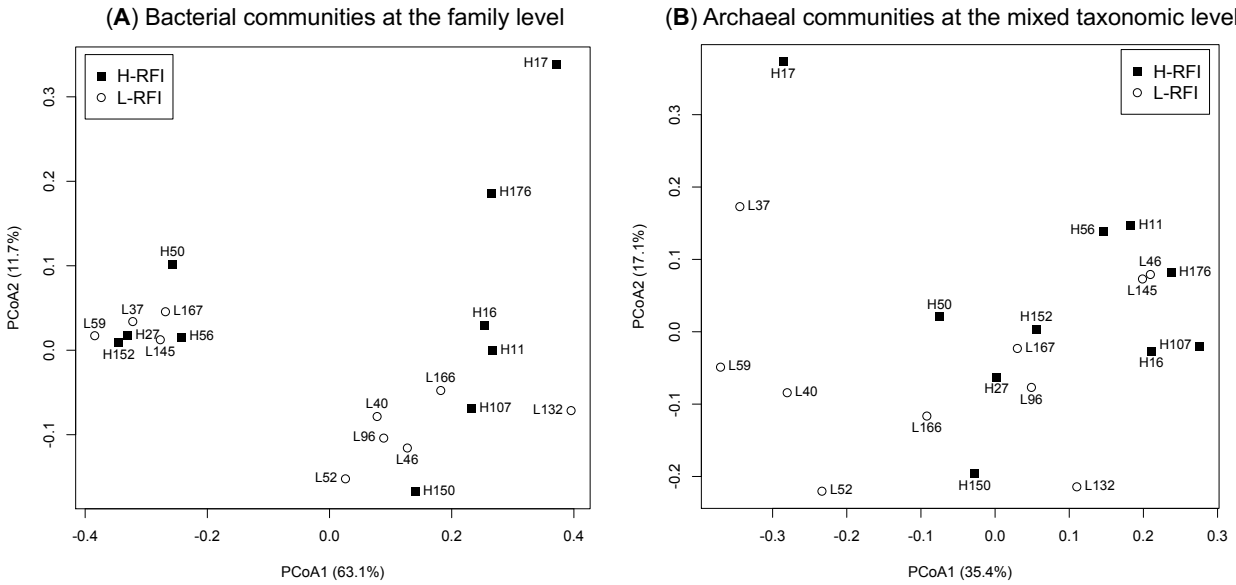


Figure 3.2 Microbial compositional profiles of H- and L-RFI animals visualized using principal coordinate analysis (PCoA). The first two PCoAs were plotted and they were calculated based on the Bray-Curtis dissimilarity matrices at the bacterial family level (**a**) and archaeal species level (**b**).

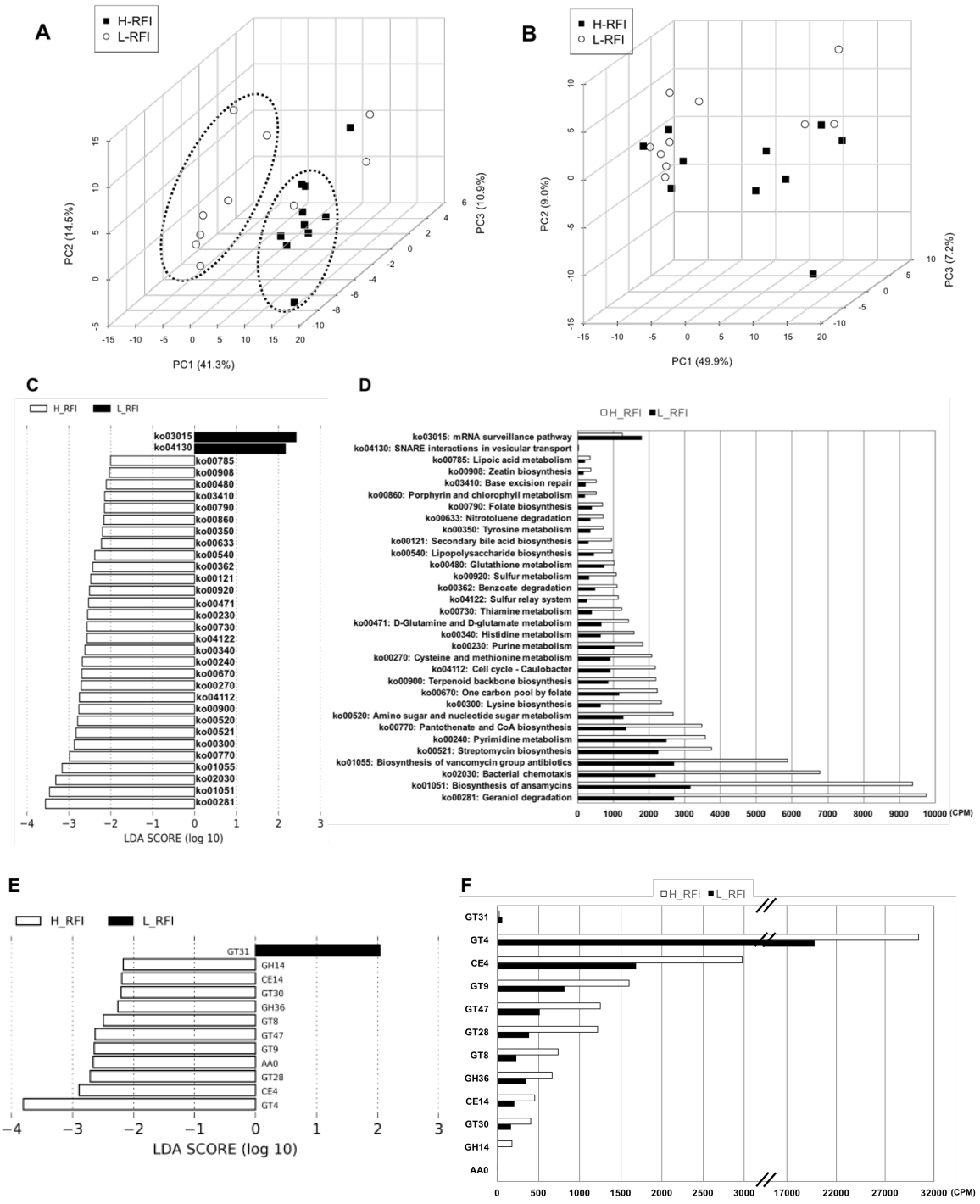


Figure 3.3 Differential rumen microbial metabolic pathways and carbohydrate-active enzymes (CAZymes) between H- and L-RFI cattle in metatranscriptomic datasets. (**a**) PCA based on microbial metabolic pathways. (**b**) PCA based on CAZymes. Histograms of Linear Discriminant Analysis (LDA) for the differential metabolic pathways (**c**) and CAZymes (**e**). Abundances of differential pathways (**d**) and CAZymes (**f**). Pathways and CAZymes with LDA score > 2 and $P < 0.05$ were considered as significantly differential features.

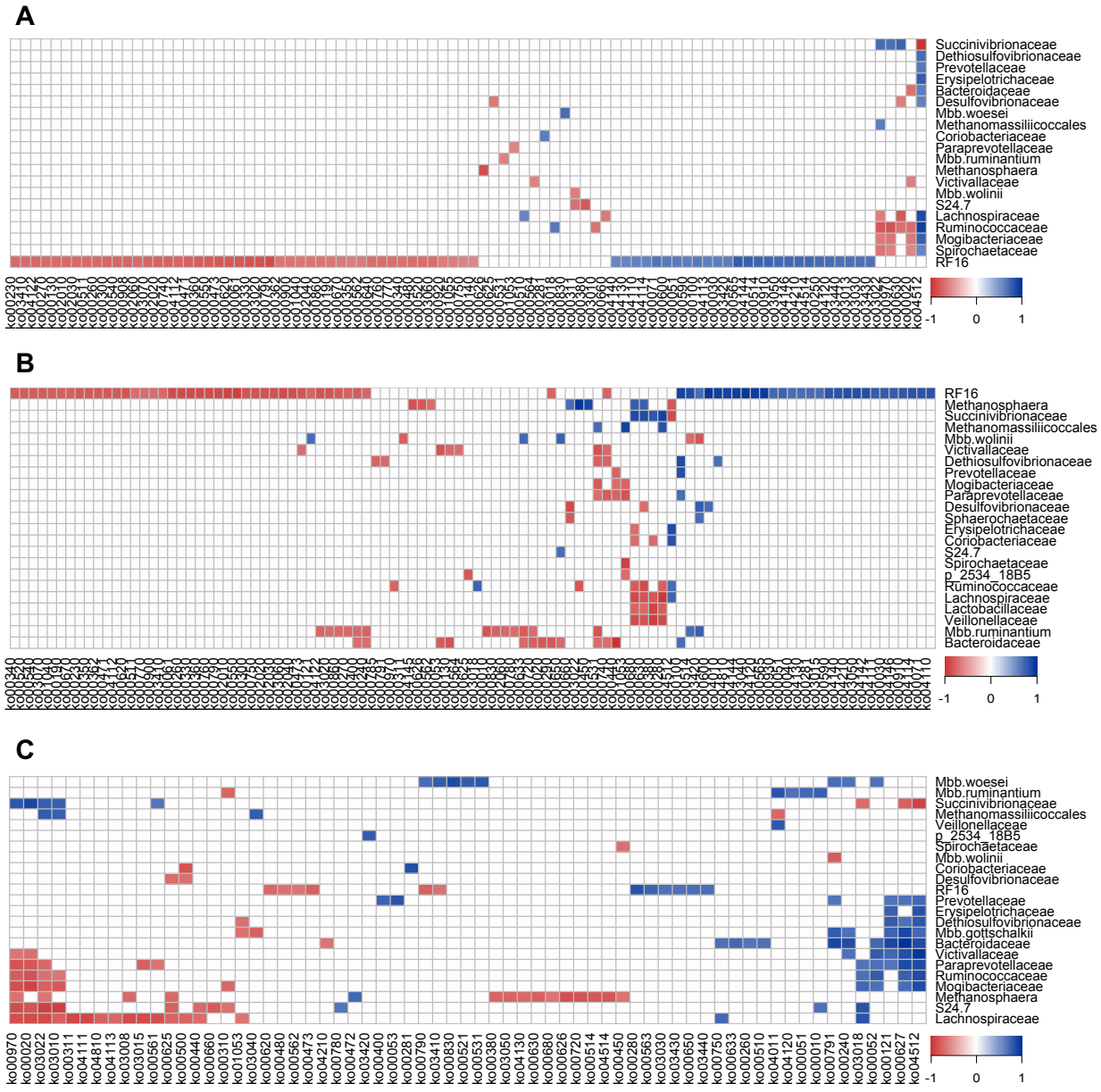


Figure 3.4 Correlation patterns showing associations between active microbial taxa and metabolic pathways. (**a**) Correlation patterns based on all 20 animals. (**b**) Correlation patterns within the H-RFI group. (**v**) Correlation patterns within the L-RFI group. Correlation analyses were conducted using Spearman's rank correlation. Only strong (correlation coefficient [ρ] > 0.6 or < -0.6) and significant ($P < 0.05$) correlations were chose to be displayed. The scale ranged from -1 (red) to 1 (blue).

Chapter 4. Metagenomics and metatranscriptomics reveal the breed effect on the rumen microbiome and its associations with feed efficiency in beef cattle

4.1 Introduction

Beef cattle provide high quality protein sources (meat) and economic profits to humans. With the increase of global human population, there are increased resource competitions (e.g., land, water, and cereal grains) between human and livestock, especially for the beef operations (Eisler *et al.*, 2014; Thornton, 2010). Improving feed efficiency of beef cattle could enhance the feed utilization ratio, and thus make them consume less feed (especially human edible cereal grains) but maintain higher or equal production performance. Meanwhile, 35 - 40% of anthropogenic CH_4 (one of greenhouse gases) emissions come from the livestock sector (Steinfeld *et al.*, 2006), which are largely produced during the rumen fermentation and manure handling in the ruminant operations (Xue *et al.*, 2014). Improving feed efficiency can decrease these negative environmental effects caused by the beef cattle operations, because cattle with high feed efficiency cattle not only emit less CH_4 (~25%), but also excrete less feces than those with low feed efficiency (Hegarty *et al.*, 2007; Nkrumah *et al.*, 2006).

The rumen microbiota consists of bacterial, archaea, fungi, ciliated protozoa, and phages (Morgavi *et al.*, 2013). They are responsible for the rumen fermentation and several studies have reported their associations with feed efficiency in beef and/or dairy cattle. However, most of these studies only focused on the taxonomic profiles. Differences in relative abundance of several rumen microbial phylotypes were reported between high and low feed efficiency individuals (Carberry *et al.*, 2012; Hernandez-Sanabria *et al.*, 2010; Jewell *et al.*, 2015; Myer *et al.*, 2015a; Zhou *et al.*, 2009); alpha-diversity indices of rumen bacterial and archaeal communities also

contributed to the variations of feed efficiency in cattle, where inefficient individuals possessed more complex and diverse microbial communities (Shabat *et al.*, 2016; Zhou *et al.*, 2009).

Nowadays, metagenomics and metatranscriptomics are becoming powerful tools to study and link rumen microbial functions to host feed efficiency. Using these approaches, two recent studies linked rumen microbial functional profiles to feed efficiency in cattle (Shabat *et al.*, 2016; Chapter 3). Both of these studies suggest that rumen microbiomes from inefficient cattle may have more diverse functional potentials and higher activities than those in efficient cattle, leading to a wider range of fermentation products, but these products may not be efficiently utilized by the host and/or even harmful to the host. To date, previous studies on the associations between the rumen microbiome and feed efficiency were conducted without the consideration of the host genetic effect. However, previous studies have suggested that rumen microbial profiles were distinguishable among hosts with different genetic backgrounds (Henderson *et al.*, 2015; Hernandez-Sanabria *et al.*, 2013). This could partially explain that association patterns between the rumen microbiome and feed efficiency showed low consistency among different studies. Therefore, a more comprehensive study is needed to understand how rumen microbiomes interact with hosts from different genetic backgrounds and how these interactions contribute to feed efficiency in beef cattle.

In the present study, rumen microbiomes of beef cattle from three different breeds with extremely high and low feed efficiency performances were explored using metagenomics, total-RNA-based metatranscriptomics and mRNA-enriched metatranscriptomics, aiming to generate more conclusive understanding on the role of the rumen microbiome in beef cattle feed efficiency. In addition, the direct comparison between mRNA-enriched and total-RNA-based metatranscriptomics for same rumen samples was conducted to provide useful information for further rumen metatranscriptomic study design.

4.2 Materials and Methods

4.2.1 Animal experiments and sample collection

Forty-eight steers were selected from a herd of 738 beef cattle that were born in 2014 and raised at the Roy Berg Kinsella Research Ranch, University of Alberta, according to their breeds and residual feed intake (RFI) ranking. These 48 steers belong to three breeds and two RFI groups (high RFI [H-RFI, > mean + 0.5 × SD] and low RFI [L-RFI, < mean - 0.5 × SD]), including two purebreds (Angus [ANG]; H-RFI, n = 8; L-RFI, n = 8) and Charolais [CHAR]; H-RFI, n = 8; L-RFI, n = 8), and one crossbred (Kinsella composite hybrid [HYB]; H-RFI, n = 8; L-RFI, n = 8). The HYB population was bred from multiple beef breeds including Angus, Charolais, Galloway, Hereford, Holstein, Brown Swiss, and Simmental as described previously (Nkrumah *et al.*, 2007b). The animal study was approved by the Animal Care and Use Committee of the University of Alberta (protocol no. AUP00000777), following the guideline of the Canadian Council on Animal Care (Olfert *et al.*, 1993). Dry matter intake (DMI) and eating frequency (times of an individual visiting the feed bunk per day) were individually recorded using the GrowSafe system (GrowSafe Systems Ltd., Airdrie, AB, Canada). RFI values were calculated based on DMI, average daily gain (ADG), metabolic weight (MWT), and backfat thickness as descried by Basarab *et al.* (2011). Steers were slaughtered before feeding at Lacombe Research Centre (Agriculture and Agri-Food Canada, Lacombe, AB, Canada). Rumen disgesta samples were collected at slaughter, snap-frozen using liquid nitrogen, and stored under -80°C until further analysis. Rumen weight was obtained after completely emptying rumen digesta and fluid using a weight balance.

4.2.2 DNA extraction and metagenome sequencing

Total genomic DNA was isolated from rumen digesta using the repeated bead beating

plus column (RBB + C) method as described by Yu and Morrison (2004). The quality and quantity of DNA was measured using a NanoDrop Spectrophotometer ND-1000 (Thermo Fisher Scientific Inc., Wilmington, DE, USA). Metagenome library was constructed using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina, San Diego, CA, USA), and the quantity of each library was evaluated using a Qubit 2.0 fluorimeter (Invitrogen, Carlsbad, CA, USA). Sequencing of metagenome libraries was conducted at the McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada) using Illumina HiSeq 2000 (100 bp paired-end sequencing of ~350 bp inserts).

4.2.3 RNA extraction and metatranscriptome sequencing

Total RNA was extracted from rumen disgesta following the procedure described in Chapter 2. The RNA yield was measured using a Qubit 2.0 fluorimeter (Invitrogen), and the RNA quality was measure using an Agilent 2200 TapeStation (Agilent Technologies, Santa Clara, CA, USA). Only samples with RNA integrity number (RIN) $\geq$ 7.0 were used to generate metatranscriptome libraries. In the current study, two types of metatranscriptome libraries were constructed: total-RNA-based metatranscriptome libraries (T-metatranscriptome) and mRNA-enriched metatranscriptome libraries (M-metatranscriptome). For the M-metatranscritpome library construction, rRNA in each sample was depleted using Ribo-Zero Gold rRNA Removal Kit (Epidemiology) (Illumina) according to the manufacturer's instruction. Total RNA and enriched mRNA was used for T- and M-metatranscriptome library construction, respectively, using TruSeq RNA Library Prep Kit v2 (Illumina). Sequencing of T- amd M-metatranscriptome libraries was conducted at the McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada) using Illumina HiSeq 2000 (100 bp paired-end sequencing of ~140 bp inserts) and 2500 (125 bp paired-end sequencing of ~140 bp inserts), respectively.

4.2.4 Analysis of metagenomes and metatranscriptomes

The quality control of each dataset was performed to trim artificial sequences (adapters), cut low quality bases (quality scores < 20), and remove short reads (< 50 bp) using Trimmomatic (version 0.35) (Bolger *et al.*, 2014). The program SortMeRNA (version 1.9) (Kopylova *et al.*, 2012) was used to extract rDNA and rRNA reads from sequencing datasets. Non-rDNA/rRNA reads were then mapped to the bovine genome (UMD 3.1) using Tophat2 (version 2.0.9) (Kim *et al.*, 2013) to remove potential host DNA and RNA contaminations. Taxonomic profiles of the active rumen microbiota were generated using 16S rRNA extracted from T-metatranscriptomes following the pipeline described in Chapter 2.

To estimate rumen microbial functional profiles, non-rDNA sequences from metagenomes (n = 48) were pooled together, and were then *de novo* assembled using Spherical program (https://github.com/thh32/Spherical). During running Spherical, Velvet was set as the assembler with the kmer size of 31, Bowtie2 was set as the aligner, and 25% of total pooled sequences were subsampled as the input for each iteration of assembly with eight iterations in total. Assembled contigs were annotated using the blastx module in DIAMOND (Buchfink *et al.*, 2015) against the UniProt database (The UniProt Consortium, 2017), and only annotations with bitscore > 40 were kept for the downstream analysis. Overlapped annotations were filtered and converted into the GFF format using the MGKit package (https://bitbucket.org/setsuna80/mgkit). Non-rDNA/rRNA sequences from metagenomes (n = 48), T-metatranscriptomes (n = 48), M-metatranscriptomes (n = 48) were individually aligned to assembled contigs using Bowtie2 (Langmead and Salzberg, 2012) and then counted using HTSeq (Anders *et al.*, 2015). Only reads mapped to contigs with eggNOG annotation information were further retrieved to calculate the abundance of genes and functional categories using MGKit (Huerta-Cepas *et al.*, 2016).

4.2.5 Statistical analysis

Values of RFI, DMI, eating frequency, and rumen weight were compared among three breeds using ANOVA, and the comparison between efficient (L-RFI) and inefficient (H-RFI) animals were conducted using *t* test within each breed separately. In the current study, only microbial taxa with a relative abundance higher than 0.01% in at least 50% of individuals within each breed were considered as being observed and used for the analysis. Bacterial compositional profiles were summarized at phylum and genus levels, and archaeal communities were summarized at the species level, respectively. Relative abundance of microbial taxa was arcsine square root transformed (Franzosa *et al.*, 2014), and then compared among breeds (using ANOVA) and between RFI groups within each breed (using *t* test). To make alpha-diversity indices (including Chao1, Shannon evenness, Simpson evenness, Shannon index, and inverse Simpson) comparable among samples, the number of sequences per sample was normalized to the lowest reads number for bacteria (n = 274,885) and archaea (n = 4,263), respectively. These indices were compared between H- and L-RFI groups within each breed using Kruskal-Wallis rank sum test. Principal-coordinate analysis (PCoA) was used to virtualize rumen microbial communities based on the Bray-Curtis dissimilarity matrices at the genus level for bacteria and at the species level for archaea.

Only functional categories and genes/transcripts with a minimum relative abundance of 0.01% in at least three samples within a dataset were considered as being detected as suggested by Franzosa *et al.* (2014). The abundance was then normalized into counts per million (cpm). To compare general microbial functional profiles among different datasets, breeds, and RFI groups, the principal component analysis (PCA) was conducted based on the auto-scaled cpm of functional categories and genes (or transcripts). Correlations between datasets were calculated using Spearman's rank correlation. Abundance of functional categories and genes (or transcripts)

were compared among sequencing datasets, breeds, and RFI groups using DESeq2 (Love *et al.*, 2014).

4.3 Results and Discussion

Residual feed intake (RFI) values were not discriminated among three breeds ($P = 0.73$), but they were significant different between L- and H-RFI animals within each breed ($P < 1.00e\text{-}5$; **Table 4.1**). After quality control, a total of 2622.07M, 2645.13M, and 3087.41M sequences were generated with an average of 54.63M, 55.11M, and 64.32M per sample for metagenomes, T-metatranscriptomes, and M-metatranscriptomes, respectively. From metagenomes/T-/M-metatranscriptomes, 99.37 ± 0.03% / 6.29 ± 0.16% / 53.34 ± 2.14% (mean ± SEM) sequences were categorized into non-rDNA/rRNA, and sequences aligned to the bovine genome were lower than 0.20% in all three datasets (**Table 4.2**).

4.3.1 Compositional profiles of the active rumen microbiota

From T-metatranscriptomes, a total of 38,610,728 sequences were enriched as bacterial V1-V3 regional 16S rRNA (804,390 ± 63,802; mean ± SEM) and 745,816 sequences as archaeal V6-V8 regional 16S rRNA (15,538 ± 1,388), which were used to generate compositional profiles of active rumen bacterial and archaeal communities. It is notable that there were only 42.15% and 64.39% bacterial and archaeal sequences falling within named genera and named species, respectively (**Supplementary Table S4.1**). The high proportion of unclassified taxa at the deep taxonomic level emphasizes that more effort, especially isolation and pure culture based studies are necessary to comprehensively characterize rumen microorganisms. In the present study, to better represent rumen microbial communities and detect the potential associations between microbial taxa and feed efficiency, unnamed and/or unclassified taxa were included in the

analysis.

In total, 15 bacterial phyla, 108 bacterial genus-level taxa, and 24 archaeal species-level taxa were identified from T-metatranscriptomes (**Supplementary Table S4.1**). Among them, 13 bacterial phyla, 66 bacterial genus-level taxa, and 16 archaeal species-level taxa were detected across all samples, confirming the idea that there is a core rumen microbiota (Henderson *et al.*, 2015). The dominant bacteria phylum was *Bacteroidetes* (26.32 ± 1.34%), followed by *Firmicutes* (25.74 ± 0.91%), *Spirochaetes* (12.81 ± 0.99%), and *Proteobacteria* (11.04 ± 1.54%). At the genus level, *Prevotella* (11.94 ± 0.49%), *Treponema* (11.25 ± 0.95%), unnamed *Succinivibrionaceae* (8.98 ± 1.50%), unclassified *Bacteroidales* (6.05 ± 0.29%), and *Fibrobacter* (6.01 ± 0.64%) were the most abundant bacterial taxa. The rumen archaeal community was dominated by *Methanobrevibacter ruminantium* (27.58 ± 1.50%), unclassified *Methanomassiliicoccaceae* (19.53 ± 1.12%), Group12 sp. ISO4-H5 (*Methanomassiliicoccaceae*-affiliated; 11.05 ± 1.20%), and *Methanobrevibacter gottschalkii* (10.22 ± 1.09%) (**Figure 4.1** and **Supplementary Table S4.1**). These results are generally comparable to previous described rumen microbial profiles (Henderson *et al.*, 2015; Kim *et al.*, 2011).

4.3.2 Breed effect on the active rumen microbiota

The distribution of detected active microbial taxa was not equal, when it was compared among three breeds (**Figure 4.1**). Although breed did not influence any alpha-diversity indices ($P > 0.05$ by Kruskal-Wallis rank-sum test; **Supplementary Table S4.1**), the principal-coordinate analysis (PCoA) showed that rumen bacterial and archaeal communities in HYB were distinct from those in ANG and CHAR (**Figure 4.2**). Comparison analysis based on the arcsine square root-transformed relative abundance revealed that around ~50% of observed microbial taxa were affected by breed, including 8 bacterial phyla (e.g., *Bacteroidetes*, *Spirochaetes*, etc.), 55 taxa at

the genus level (e.g., *Prevotella*, *Treponema*, etc.), and 10 species-level archaeal taxa (e.g., *Methanomassiliicoccaceae*-affiliated Group12 sp. ISO4-H5, unclassified *Methanobrevibacter*, etc.; **Supplementary Table S4.1**).

Several biological factors potentially contribute to the rumen microbiota variations among breeds. Firstly, we observed significantly different eating frequencies among three breeds: HYB showed lower eat frequency (29.73 ± 1.99 time/day) than that of ANG and CHAR (37.63 ± 1.61 and 36.59 ± 1.56 time/day, respectively; P = 6.30e-03) (**Table 4.1**). Because salivation is enhanced during eating than resting (Beauchemin *et al.*, 2008), lower eating frequencies may lead to lower amounts of saliva produced in HYB, which consequently results in the shift of rumen pH and thus influences the rumen microbiota (Hernandez et al., 2014). Meanwhile, ANG and CHAR had higher feed intake (DMI; 10.73 ± 0.27 and 10.33 ± 0.28 kg/day, respectively) than HYB (9.27 ± 0.28 kg/day; P = 3.23e-03) (**Table 4.1**). It is known that the growth of rumen microbiota is positively correlated with feed intake due to more available substrates and nutrients for the microbial growth (Pathak, 2008; Singh *et al.*, 1977). Furthermore, we also observed different rumen sizes due to the breed effect (P = 1.36e-02) (**Table 4.1**). Both feed intake and rumen size have impact on the rumen passage rate (Colucci *et al.*, 1982). And the rumen passage rate could then affect the rumen microbial growth (Martinez *et al.*, 2009; Shriver *et al.*, 1986), because it is associated with the microbial energy flux (maintenance vs. growth) and microbial generation times (Firkins, 1996; Sniffen and Robinson, 1987). In addition, it has been suggested that the increased rumen passage rate and washout decreased the abundance of rumen methanogens (Janssen, 2010), which was further confirmed in a recent study that revealed low CH_4 yield sheep had smaller rumen size and shorter rumen retention time (Goopy *et al.*, 2014). Although effects of those biological factors on the rumen microbial growth/abundance have been widely reported as discussed above, how those factors contribute to the variations of microbial

composition have not been well described. Therefore, further studies to link those biological factors to microbial compositional profiles are needed, which could help us better understand the breed effect on the rumen microbiota as observed in this study.

4.3.3 Breed effect on differential microbial taxonomic features between RFI groups

As breed-associated differences were observed for ~50% of bacteria and archaeal taxa, analyses of relationship between rumen microbial features and feed efficiency were performed under each breed. Relative abundance of *Firmicutes* (L-RFI: 28.56 ± 1.82% vs. H-RFI: 22.45 ± 2.14%; $P = 0.042$) and *Chloroflexi* (L-RFI: 0.05 ± 0.01% vs. H-RFI: 0.03 ± 0.01%; $P = 0.046$) were different between H- and L-RFI CHAR steers, while no bacterial phylum had different abundance between RFI groups in HYB and ANG. When the analysis was conducted at the bacterial genus level, 22 (e.g., unnamed *Bacteroidales*, *Butyrivibrio*, etc.), one (unnamed RF16), and 16 genus-level taxa (e.g., unclassified *Clostridiales*, unnamed *Ruminococcaceae*, etc.) were differentially abundant between H- and L-RFI steers in HYB, ANG, and CHAR, respectively ($P < 0.05$; **Table 4.3**). For archaea, differences in abundance of *Methanobrevibacter smithii* and four taxa (unclassified *Methanomassiliicoccaceae*, unclassified *Methanobrevibacter*, unclassified Group11, and *Methanomethylophilus alvus*) were detected between H- and L-RFI steers ($P < 0.05$) in HYB and CHAR, respectively, but no differential archaeal taxon was detected between RFI groups in ANG (**Table 4.3**). Meanwhile, H- and L-RFI HYB steers differed in Shannon index of bacterial communities ($P = 0.04$). For CHAR steers, two RFI groups had significantly different inverse Simpson ($P = 0.03$) and Simpson evenness ($P = 0.03$) of archaeal communities, as well as Shannon evenness of bacteria communities ($P = 0.03$) (**Table 4.4**).

Differential taxonomic features between H- and L-RFI groups were not consistent among three breeds, except for four differential bacterial genus-level taxa in HYB and CHAR (*Blautia*,

unclassified *Clostridia*, unnamed *Mogibacteriaceae*, and unnamed R4-45B). Although these bacterial taxa were low abundant in the rumen (< 0.5%), it is notable that they all showed higher abundance in L-RFI animals than in H-RFI individuals in both HYB and CHAR (**Table 4.3**). *Blautia* members are ubiquitously distributed in mammal gut with low abundance (Eren *et al.*, 2015). They have been reported to provide energy to hosts from the fermentation of polysaccharides that other microbial taxa cannot (Biddle *et al.*, 2013), and thus higher abundance of *Blautia* may extend the rumen metabolic capacity for steers with high feed efficiency. In addition, members of *Blautia*, such as *Blautia hydrogenotrophica*, have the capacity to consume H_2 and produce acetate through acetogenesis (Rey *et al.*, 2010). Therefore, the increased abundance of *Blautia* indicates higher acetogenesis in L-RFI animals, which possibly compete with rumen methanogens. Therefore, more acetates rather than CH_4 could be generated during removing H_2 from the rumen in L-RFI individuals, leading to lower energy waste. *Mogibacteriaceae*-affiliated unnamed genus has already been reported to be associated with feed efficiency in beef cattle with multiple genetic backgrounds (Myer *et al.*, 2015a), but scarce information is available to define its functions in the gut. Abundance of members in this family were negatively correlated with body mass index (BMI) in human (Goodrich *et al.*, 2014; Oki *et al.*, 2016). Therefore, the higher abundance of *Mogibacteriaceae* in L-RFI individuals may correspond to a leaner body type, which may further correspond to a higher protein deposition in individuals with high feed efficiency.

All 48 steers involved in this study received identical diet and were raised under the same environment, but rumen microbial communities were distinguishable among different breeds and there were unique differential taxonomic features between RFI groups within each breed. Therefore, it is reasonable to speculate that there are several rumen microorganisms belonging to different taxonomic groups sharing similar ecologic niches among different hosts. In other words,

different microbial phylotypes may utilize similar substrates and produce similar products in the rumen. Indeed, previous studies demonstrated that functional profiles of the microbiome were more conserved than the taxonomic composition at a certain body site in human (Franzosa *et al.*, 2014; Human Microbiome Project Consortium, 2012). In ruminants, Taxis *et al.* (2015) suggested that rumen microbiomes were dissimilar at the taxonomic level but considerably more similar at the metabolic functional level. Recently, Roehe *et al.* (2016) and findings in Chapter 3 further revealed that methane emissions and RFI were more associated with rumen microbial functional features rather than taxonomic features. Collectively, merely analyzing rumen microbial communities may be not powerful enough to detect, even mask actual linkages between rumen microbiomes and feed efficiency. Therefore, it is necessary to further investigate how rumen microbial functional features contribute to the variations of feed efficiency.

4.3.4 Functional profiles of the rumen microbiome at DNA and RNA levels

The *de novo* assembly of non-rDNA metagenome reads generated a total of 57,696,422 contigs, with an average length of 144 bp (max 135,846 bp) and a N50 length of 140 bp. After filtering overlapped annotations, 20,314,713 contigs (35.21%) were successfully annotated based on the UniProt database. An average of 78.47 ± 0.26%, 66.85 ± 0.65%, and 54.43 ± 1.16% of sequences from metagenomes, T-metatranscriptomes, and M-metatranscriptomes could be mapped back to assembled contigs, respectively, suggesting the assembly was successful. In total, 3,589,489 contigs were annotated with eggNOG information, and only reads mapped to these contigs were retrieved to estimate functional profiles. Detailed sequencing datasets are listed in **Table 4.2**.

In total, 23 eggNOG functional categories were observed through the functional analysis at both DNA and RNA levels. For metagenome reads, 10.43%, 8.15%, and 8.10% of them were

involved in "*Replication, recombination and repair*", "*Amino acid transport and metabolism*", and "*Carbohydrate transport and metabolism*", respectively, and 20.07% were poorly characterized. For both T- and M-metatranscriptomes, "*Carbohydrate transport and metabolism*" was the most active functional category (13.96% and 13.78% in T- and M-metatranscriptomes, respectively), followed by the functional category of "*Translation, ribosomal structure and biogenesis*" (9.22% and 8.66% in T- and M-metatranscriptomes, respectively) (**Figure 4.3**). Accordingly, these results suggest that microbiomes may possess large genetic potentials to survive and grow in the rumen, but they were mainly conducting fermentation to degrade feed at the point when digesta samples were collected in this study.

4.3.4.1 Comparison between metagenomes and metatranscriptomes

The principle component analysis (PCA) based on eggNOG functional categories showed clear separation between metagenome and metatranscriptome functional profiles (**Figure 4.4a**). Compared with T- and M-metatranscriptomes, metagenomes from rumen digesta samples were more closely clustered together and thus more conserved among individuals (**Figure 4.4a**), suggesting that inter-individual functional variations at the RNA level were higher than those at the DNA level. Therefore, rumen microbiomes from different individuals may have similar functional genetic potentials (at the DNA level), while their actual functional activities (at the RNA level) are noticeably more variable, which are similar to the findings of human gut microbiomes (Franzosa *et al.*, 2014; Nayfach *et al.*, 2015). To date, most existing associations between the rumen microbiome and host phenotypes (e.g., feed efficiency, methane emissions, and so on) are built on at the DNA level (Hernandez-Sanabria *et al.*, 2010; Shabat *et al.*, 2016; Wallace *et al.*, 2015; Zhou *et al.*, 2010). However, Shi *et al.* (2014) reported that differences of rumen microbial gene expression profiles, rather than genomic profiles were associated with the

variations of CH_4 emissions in sheep. Collectively, host phenotypic performances may be more associated with rumen microbial activities (RNA) than functional genetic potentials (DNA), and thus analysis at the RNA level is a more meaningful approach to link the rumen microbiome to host performances.

Several functional categories were abundant at the DNA level but down-expressed at the RNA level, such as "*Replication, recombination and repair*" (~2 folds), "*Extracellular structures*" (~3 folds), "*Defense mechanisms*" (~3 folds), and so on (**Figure 4.3** and **Figure 4.4b** and **4.4c**). These functional categories were less active in the rumen of steers used in this study, but they represent large functional potentials under environmental challenges. For instance, "*Defense mechanisms*" were down-regulated in both T- and M-metatranscriptomes, however, their high abundance at the DNA level suggest that they could be activated when rumen microbiomes respond to adverse conditions, such as dietary change, abrupt pH drop, and so on. Meanwhile, some function categories, including "*Carbohydrate transport and metabolism*", "*Translation, ribosomal structure and biogenesis*", "*Cell motility*", "*Cytoskeleton*", etc., were overexpressed at the RNA level beyond their DNA abundance (2 - 6 folds). The highest activity of "*Carbohydrate transport and metabolism*" among all rumen functions is consistent with the rumen metatranscriptome outcomes reported in Chapter 3, indicating most of active microorganisms are fermenting feed carbohydrates (e.g., cellulosic plant materials, starch, etc.) when digesta samples were collected.

Although general functional profiles were different between DNA and RNA levels, strong correlations were detected between metagenomes and metatranscriptomes (Spearman's $r = 0.91$, $P = 1.88e\text{-}6$ between metagenomes and T-metatranscriptomes; $r = 0.92$, $P = 1.69e\text{-}6$ between metagenomes and M-metatranscriptomes; **Figure 4.4b** and **4.4c**), which is line with the correlation pattern between human gut metagenomes and metatranscriptomes (Franzosa *et al.*,

2014). Through the linear regression estimation, metagenomes could explain 57.57% (P = 2.61e-06) and 60.81% (P = 6.67e-06) of variations in T- and M-metatranscriptomes, respectively. These strong correlations suggest that gene expressions of microbiomes highly depend on their gene abundance, although there are other factors (such as environmental factors and post transcriptional regulation) contributing to microbial gene expression variations in the rumen.

4.3.4.2 Comparison between M- and T-metatranscriptomes

The mRNA enrichment step significantly removed rRNA from total RNA. There was 93.71 ± 0.16% rRNA in T-metatranscriptomes but only 46.66 ± 2.14% rRNA in M-metatranscriptomes (P = 7.36e-26; Paired Sample t Test; **Table 4.2**), indicating a successful rRNA removal using the Ribo-Zero Gold rRNA Removal Kit. It is worth mentioning that the majority of remained rRNA in M-metatranscriptomes was classified as eukaryotic 28S rRNA (34.71 ± 0.16%), because this kit is mainly designed to hybridize and remove prokaryotic rRNA rather than eukaryotic rRNA. A higher proportion of T-metatranscriptome reads could be mapped back to assembled metagenome contigs than M-metatranscriptome reads (66.85 ± 0.65% versus 54.43 ± 1.16%, P = 6.42e-18; Paired Sample t Test). This suggests that T-metatranscriptomes are more similar with metagenomes, while M-metatranscriptomes may capture extra expressed genes with low abundance. In addition, there were more annotated mRNA reads in M- than in T-metatranscriptomes (412,875 ± 30,166 vs 23,590 ± 1,494, P = 1.16e-17; Paired Sample t Test), supporting that the mRNA enrichment indeed increase the outcomes of mRNA.

According to PCA, overall functional profiles did not show clear difference between T- and M-metatranscriptomes (**Figure 4.5a** and **4.5b**). At the same time, strong correlations were detected between T- and M-metatranscriptome based on both functional categories (Spearman's r = 1.00, P = 3.61e-7) and expressed genes (Spearman's r = 0.84, P < 2.20e-16) (**Figure 4.5c** and

4.5d). The linear regression analysis based on functional categories and expressed genes gave R^2 value of 1.00 (P < 2.2e-16) and 0.94 (P < 2.2e-16), respectively, when compared T- with M-metatranscriptomes, confirming that T- and M-metatranscriptomes were highly similar to each other. When the cluster analysis was performed within each breed, T- and M-metatranscriptomes from the same sample were normally similar to each other (between-method variations < between-subject variations) except for a few samples (**Figure 4.5e**). Previous studies that compared human fecal and seawater metatranscriptomes between mRNA-enriched and total-RNA-based libraries also reported similar results (Giannoukos *et al.*, 2012; Stewart *et al.*, 2010). However, when abundance of each functional category was compared between T- and M-metatranscriptomes using the DESeq2 analysis, it identified 10 differential abundant functional categories (P < 0.05), even though their folder changes were inconspicuous (from -1.32 to 1.06; **Figure 4.5c**). At the same time, the DESeq2 analysis revealed that 2,050 genes had different abundance between T- and M-metatranscriptomes (FDR < 0.05), and most of them were underestimated in M-metatranscriptomes (**Figure 4.5d**). In line with our results, Tveit *et al.* (2014) also found that the mRNA enrichment step had a significant impact on the abundance estimation of each transcript. Because the mRNA enrichment step in this study was based on the rRNA depletion strategy instead of PCR amplifying mRNA, it is reasonable to speculate that the underestimation of many expressed genes in M-metatranscriptomes may be caused by the mRNA degradation during the extended sample processing time.

To date, the mRNA enrichment is normally conducted prior to the metatranscriptome library construction to increase the resolution of mRNA and capture more transcripts for the rumen microbiome (AlZahal *et al.*, 2017; Kamke *et al.*, 2016; Shi *et al.*, 2014). But some other rumen metatranscriptomic studies were conducted based on the sequencing of total RNA without the mRNA enrichment, which also successfully generate functional profiles for rumen

microbiomes (Poulsen *et al.*, 2013; Chapter 3). According to our results, although applying the mRNA enrichment could increase the sequencing depth of mRNA and enhance the resolution of metatranscriptomics on the functional analysis, it may bring about biases for the estimation of gene expression levels. In contrast, total-RNA-based metatranscriptomcics not only generates similar functional profiles as mRNA-enriched metatranscriptomics, but also can be used for the taxonomic identification (Chapter 2). Considering the rapid reduction of NGS costs, plus our current findings descripted above, total RNA sequencing rather than enriched-mRNA sequencing is more recommended, when researchers plan to globally screen the compositional and functional characteristics of the rumen microbiome. However, if expression levels of specific genes and/or metabolic pathways need to be explored, the mRNA enrichment is still necessary to be applied for the enhanced resolution of metatranscriptomics.

4.3.5 Breed effect on differential microbial functions between H- and L-RFI steers

Rumen microbiomes from three breeds showed distinguishable functional profiles at the DNA level, especially that microbiomes from HYB were separated with those from ANG and CHAR (**Figure 4.6a**). However, at the transcriptomic level (in both T- and M-metatranscriptomes), differences among three breeds were not obvious (**Figure 4.6b** and **4.6c**). These indicate that host genetics may influence functional genetic potentials of rumen microbiomes, but their actual activities may be less impacted by host genetics and are more driven by environmental factors. And the lack of separation of metatranscirpomes among breeds may due to the same diet and environmental conditions for these steers.

In ANG, RFI had no effect on functional category identified from metagenomes and T-metatranscriptomes, while "*RNA processing and modification*" showed higher abundance in M-metatranscriptomes of L-RFI animals than that of H-RFI ones (P = 0.021). For CHAR, two

functional categories "*Cell cycle control, cell division, chromosome partitioning*", and "*Secondary metabolites biosynthesis, transport and catabolism*" were more abundant in H-RFI animals than in L-RFI animals at the genomic level ($P = 0.008$ and 0.033, respectively). In T- and M-metatranscritpomes, four and two functional categories were differentially abundant between RFI groups, respectively. Especially, "*Translation, ribosomal structure and biogenesis*" and "*Transcription*" had higher expression levels in H-RFI animals from both T- and M-metatranscritpomes ($P < 0.05$; **Table 4.5**). For HYB steers, "*Intracellular trafficking, secretion, and vesicular transport*" was higher abundant in H-RFI steers than in L-RFI steers at the DNA level ($P = 0.014$). "*Cell motility*" was more abundant at the transcriptomic level revealed from both T- and M-metatranscritpomes ($P = 0.044$ and 0.013, respectively). "*Nucleotide transport and metabolism*" and "*Cytoskeleton*" only showed different abundance in T-metatranscritpomes ($P = 0.010$ and 0.036, respectively).

Comparison analysis of metagenomes revealed 932 genes with different abundance between H- and L-RFI animals from metagenomes: 591 genes in CHAR, 216 genes in HYB, and one gene in ANG, with 124 genes overlapped in both CHAR and HYB. When compared T-metatranscriptomes, there were 39 differentially expressed genes between RFI groups (29 in HYB, ten in CHAR, and none in ANG). From the comparison of M-metatranscriptmes, RFI had effects on 14 expressed genes (12 in HYB and two in CHAR) (**Figure 4.7a-c**). It is notable that only three differential genes were detected between H- and L-RFI steers at both DNA and RNA levels: two were overlapped between metagenomes and M-metatranscriptomes (genes coding 2,3-bisphosphoglycerate-independent phosphoglycerate mutase and coding fumarate reductase/succinate dehydrogenase flavoprotein domain protein), and one was overlapped between T- and M-metatranscriptomes (gene coding phosphoketolase) (**Figure 4.7d**).

Recent studies suggest that rumen microbiomes of H-RFI animals have higher and more diverse functional genetic potentials and activities than those of L-RFI individuals (Shabat *et al.*, 2016; Chapter 3). Those findings are further confirmed in the present study, especially at the gene/transcript level, that most of differential microbial genes/transcripts between RFI groups were enriched in H-RFI steers than in L-RFI ones (**Figure 4.7a-c**). Hence, it can be speculated that rumen microorganisms of inefficient individuals could ferment a wider range of substrates and generate more products. However, these products may be harmful to the host or exceed the absorbing capacity of the host, which could be regulated by host genetics and lead to low feed efficiency. Conversely, efficient cattle have relatively simpler rumen microbial functions and activities, generating more relevant products that could be efficiently absorbed by the host.

In the present study, although some microbial genes were differentially abundant between RFI groups in both CHAR and HYB metagenomes (**Figure 4.7a**), no functional category (at neither DNA nor RNA levels) or expressed gene was universally different between H- and L-RFI steers among three breeds (**Table 4.5** and **Figure 4.7b** and **4.7c**). These suggest that there are different rumen microbiome-host interaction patterns to determine the feed efficiency performance in beef cattle. For example, from all three sequencing datasets, we merely observed a handful of differential microbial features (at both compositional and functional levels) between H- and L-RFI steers in ANG, suggesting that rumen microbiomes in ANG may not or only slightly contribute to RFI variations. In contrary, many compositional and functional features of rumen microbiomes in HYB and CHAR were associated with host RFI performances. Considering different genetic backgrounds (different genotypes) among three breeds, further studies to explore interaction patterns between the rumen microbiome and host genotypes are needed to better understand how these interactions affect feed efficiency in beef cattle.

4.4 Conclusions

Taxonomic analysis using total-RNA-based metatranscriptomics revealed distinguishable active rumen microbiota among three breeds. Meanwhile, functional genetic potentials were also affected by the host genetic background according to metagenomics. These breed-associated differences represent potential superiorities of each breed, which could be further applied to manipulate the rumen microbiota through selection and breeding. However, actual activities of rumen microbiomes were less impacted by host genetics but more sensitively respond to environmental factors. In addition, rumen microorganisms possess large functional genetic potentials to survive and proliferate, with high activities at feed degradation and fermentation for rumen digesta samples collected in this study. Therefore, to better associate rumen microorganisms with host performances, metatranscriptomics is a more powerful approach. Comparison of total-RNA-based and mRNA-enriched metatranscriptomes suggested that mRNA-enriched metatranscirptomics should be used to study expression levels of specific genes and/or metabolic pathways, while total-RNA-based metatranscriptomics can be applied to generate both compositional and functional profiles simultaneously. Furthermore, several differential microbial features between RFI groups were detected within each breed, including active bacterial and archaeal taxa, alpha-diversity indices of microbial communities, functional categories and genes (at both DNA and RNA levels). These results extend our understanding on the associations between the rumen microbiome and feed efficiency at multiple genetic levels in diverse beef cattle breeds. Most of differential microbial features between H- and L-RFI steers were distinct among three breeds, suggesting there are host and microbiome interactions in the rumen contributing to feed efficiency variations. This knowledge could help us more accurately predict feed efficiency using rumen microbial features and better define approaches to manipulate the rumen microbiome to improve feed efficiency of cattle from various breeds.

4.5 References

AlZahal O, Li F, Guan LL, Walker ND, McBride BW (2017). Factors influencing ruminal bacterial community diversity and composition and microbial fibrolytic enzyme abundance in lactating dairy cows with a focus on the role of active dry yeast. *J Dairy Sci*.

Anders S, Pyl PT, Huber W (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:** 166-169.

Basarab JA, Colazo MG, Ambrose DJ, Novak S, McCartney D, Baron VS (2011). Residual feed intake adjusted for backfat thickness and feeding frequency is independent of fertility in beef heifers. *Can J Anim Sci* **91:** 573-584.

Beauchemin KA, Eriksen L, Norgaard P, Rode LM (2008). Short communication: salivary secretion during meals in lactating dairy cattle. *J Dairy Sci* **91:** 2077-2081.

Biddle A, Stewart L, Blanchard J, Leschine S (2013). Untangling the genetic basis of fibrolytic specialization by lachnospiraceae and ruminococcaceae in diverse gut communities. *Diversity* **5:** 627-640.

Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114-2120.

Buchfink B, Xie C, Huson DH (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12:** 59-60.

Carberry CA, Kenny DA, Han S, McCabe MS, Waters SM (2012). Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. *Appl Environ Microbiol* **78:** 4949-4958.

Colucci PE, Chase LE, Van Soest PJ (1982). Feed Intake, Apparent Diet Digestibility, and Rate of Particulate Passage in Dairy Cattle. *J Dairy Sc* **65:** 1445-1456.

Eisler MC, Lee MR, Tarlton JF, Martin GB, Beddington J, Dungait JA *et al* (2014). Agriculture: Steps to sustainable livestock. *Nature* **507:** 32-34.

Eren AM, Sogin ML, Morrison HG, Vineis JH, Fisher JC, Newton RJ *et al* (2015). A single genus in the gut microbiome reflects host preference and specificity. *Isme j* **9:** 90-100.

Firkins JL (1996). Maximizing microbial protein synthesis in the rumen. *J Nutr* **126:** 1347s-1354s.

Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM *et al* (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* **111:** E2329-2338.

Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ *et al* (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol* **13:** R23.

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R *et al* (2014). Human genetics shape the gut microbiome. *Cell* **159:** 789-799.

Goopy JP, Donaldson A, Hegarty R, Vercoe PE, Haynes F, Barnett M *et al* (2014). Low-methane yield sheep have smaller rumens and shorter rumen retention time. *Br J Nutr* **111:** 578-585.

Hegarty RS, Goopy JP, Herd RM, McCorkell B (2007). Cattle selected for lower residual feed intake have reduced daily methane production. *J Anim Sci* **85:** 1479-1486.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH (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.

Hernandez-Sanabria E, Guan LL, Goonewardene LA, Li M, Mujibi DF, Stothard P *et al* (2010). Correlation of particular bacterial PCR-denaturing gradient gel electrophoresis patterns

with bovine ruminal fermentation parameters and feed efficiency traits. *Appl Environ Microbiol* **76:** 6338-6350.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Zhou M, Moore SS, Guan LL (2013). Influence of sire breed on the interplay among rumen microbial populations inhabiting the rumen liquid of the progeny in beef cattle. *PLoS One* **8:** e58461.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC *et al* (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44:** D286-293.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486:** 207-214.

Janssen PH (2010). Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. *Animal Feed Science and Technology* **160:** 1-22.

Jewell KA, McCormick CA, Odt CL, Weimer PJ, Suen G (2015). Ruminal Bacterial Community Composition in Dairy Cows Is Dynamic over the Course of Two Lactations and Correlates with Feed Efficiency. *Appl Environ Microbiol* **81:** 4697-4710.

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:** 56.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14:** R36.

Kim M, Morrison M, Yu Z (2011). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol Ecol* **76:** 49-63.

Kopylova E, Noe L, Touzet H (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28:** 3211-3217.

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357-359.

Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.

Martinez ME, Ranilla MJ, Ramos S, Tejido ML, Carro MD (2009). Effects of dilution rate and retention time of concentrate on efficiency of microbial growth, methane production, and ruminal fermentation in Rusitec fermenters. *J Dairy Sci* **92:** 3930-3938.

Morgavi DP, Kelly WJ, Janssen PH, Attwood GT (2013). Rumen microbial (meta)genomics and its application to ruminant production. *Animal* **7 Suppl 1:** 184-201.

Myer PR, Smith TP, Wells JE, Kuehn LA, Freetly HC (2015). Rumen microbiome from steers differing in feed efficiency. *PLoS One* **10:** e0129174.

Nayfach S, Bradley PH, Wyman SK, Laurent TJ, Williams A, Eisen JA *et al* (2015). Automated and Accurate Estimation of Gene Family Abundance from Shotgun Metagenomes. *PLoS Comput Biol* **11:** e1004573.

Nkrumah JD, Okine EK, Mathison GW, Schmid K, Li C, Basarab JA *et al* (2006). Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle. *J Anim Sci* **84:** 145-153.

Nkrumah JD, Crews DH, Jr., Basarab JA, Price MA, Okine EK, Wang Z *et al* (2007). Genetic and phenotypic relationships of feeding behavior and temperament with performance, feed efficiency, ultrasound, and carcass merit of beef cattle. *J Anim Sci* **85:** 2382-2390.

Oki K, Toyama M, Banno T, Chonan O, Benno Y, Watanabe K (2016). Comprehensive analysis of the fecal microbiota of healthy Japanese adults reveals a new bacterial lineage

associated with a phenotype characterized by a high frequency of bowel movements and a lean body type. *BMC Microbiol* **16:** 284.

Olfert ED, Cross BM, McWilliams AA (1993). Guide to the care and use of experimental steers. *Canadian Council on Animal Care, Ottawa, ON, Canada*.

Pathak AK (2008). Various factors affecting microbial protein synthesis in the rumen. *Vet World* **1:** 186-189.

Poulsen M, Schwab C, Jensen BB, Engberg RM, Spang A, Canibe N *et al* (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat Commun* **4:** 1428.

Rey FE, Faith JJ, Bain J, Muehlbauer MJ, Stevens RD, Newgard CB *et al* (2010). Dissecting the in vivo metabolic potential of two human gut acetogens. *J Biol Chem* **285:** 22082-22090.

Roehe R, Dewhurst RJ, Duthie CA, Rooke JA, McKain N, Ross DW *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 Genet* **12:** e1005846.

Shabat SK, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Berg Miller ME *et al* (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *Isme j*.

Shi W, Moon CD, Leahy SC, Kang D, Froula J, Kittelmann S *et al* (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res* **24:** 1517-1525.

Shriver BJ, Hoover WH, Sargent JP, Crawford RJ, Jr., Thayne WV (1986). Fermentation of a High Concentrate Diet as Affected by Ruminal pH and Digesta Flow. *J Dairy Sci* **69:** 413-419.

Singh UB, Verma DN, Varma A, Ranjhan SK (1977). The relationship between rumen bacterial growth, intake of dry matter, digestible organic matter and volatile fatty acid production in buffalo (Bos bubalis) calves. *Br J Nutr* **38:** 335-340.

Sniffen CJ, Robinson PH (1987). Protein and fiber digestion, passage, and utilization in lactating cows. Microbial growth and flow as influenced by dietary manipulations. *J Dairy Sci* **70:** 425-441.

Steinfeld H, Gerber P, Wassenaar TD, Castel V, de Haan C (2006). *Livestock's long shadow: environmental issues and options*. Food & Agriculture Org.

Stewart FJ, Ottesen EA, DeLong EF (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *Isme j* **4:** 896-907.

Taxis TM, Wolff S, Gregg SJ, Minton NO, Zhang C, Dai J *et al* (2015). The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Nucleic Acids Res* **43:** 9600-9612.

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45:** D158-d169.

Thornton PK (2010). Livestock production: recent trends, future prospects. *Philos Trans R Soc Lond B Biol Sci* **365:** 2853-2867.

Tveit AT, Urich T, Svenning MM (2014). Metatranscriptomic analysis of arctic peat soil microbiota. *Appl Environ Microbiol* **80:** 5761-5772.

Wallace RJ, Rooke JA, McKain N, Duthie CA, Hyslop JJ, Ross DW *et al* (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* **16:** 839.

Xue B, Wang LZ, Yan T (2014). Methane emission inventories for enteric fermentation and manure management of yak, buffalo and dairy and beef cattle in China from 1988 to 2009. *Agriculture, Ecosystems and Environment* **195:** 202-210.

Yu Z, Morrison M (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36:** 808-812.

Zhou M, Hernandez-Sanabria E, Le LG (2009). Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Appl Environ Microbiol* **75:** 6524-6533.

Zhou M, Hernandez-Sanabria E, Guan LL (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. *Appl Environ Microbiol* **76:** 3776-3786.

4.6 Tables

Table 4.1 Phenotypes of beef cattle with three breeds used for this study

		Angus (ANG) (n = 16)	Charolais (CHAR) (n = 16)	Crossbred (HYB) (n = 16)	*P* value[1]
		(mean ± SEM)			
Residual feed intake (RFI; kg/day)	Overall	0.15 ± 0.21	0.06 ± 0.24	-0.10 ± 0.24	0.73
	L-RFI (n = 8)	-0.58 ± 0.10	-0.81 ± 0.10	-0.96 ± 0.10	
	H-RFI (n = 8)	0.88 ± 0.15	0.92 ± 0.12	0.76 ± 0.13	
	P value[2]	1.46E-06	1.87E-08	6.43E-08	
Dry matter intake (DMI; kg/day)	Overall	10.73 ± 0.27[a]	10.33 ± 0.28[a]	9.27 ± 0.28[b]	3.23E-03
	L-RFI (n = 8)	10.19 ± 0.34	9.49 ± 0.24	8.98 ± 0.34	
	H-RFI (n = 8)	11.35 ± 0.31	11.17 ± 0.29	9.67 ± 0.43	
	P value[2]	0.03	5.80E-04	0.23	
Eating frequency (n/day)	Overall	37.63 ± 1.61[a]	36.59 ± 1.56[a]	29.73 ± 1.99[b]	6.30E-03
	L-RFI (n = 8)	41.25 ± 1.53	35.56 ± 2.65	29.94 ± 2.02	
	H-RFI (n = 8)	33.49 ± 2.13	37.63 ± 1.77	29.44 ± 4.18	
	P value[2]	9.90E-03	0.53	0.91	
Rumen weight (kg)	Overall	13.02 ± 0.59[a]	11.29 ± 0.29[b]	NA	6.30E-03
	L-RFI (n = 8)	12.42 ± 0.80	11.15 ± 0.38	NA	
	H-RFI (n = 8)	13.62 ± 0.87	11.43 ± 0.47	NA	
	P value[2]	0.33	0.65	NA	

1. *P* values among three breeds were calculated using ANOVA, and values with different superscripts were significantly different ($P < 0.05$).
2. *P* values between H- and L-RFI groups were obtained using *t* test within each breed.

Table 4.2 Summary of sequencing data for metagenomes and metatranscriptomes

	Metagenome (n = 48)	T-metatranscriptome (n = 48) (mean ± SEM)	M-metatranscriptome (n = 48)
After QC[1]	54.63 ± 1.42 M	64.32 ± 0.74 M	55.11 ± 1.90 M
non-rDNAs/rRNAs	99.37 ± 0.03%	6.29 ± 0.16%	53.34 ± 2.14%
rDNAs/rRNAs	0.63 ± 0.03%	93.71 ± 0.16%	46.66 ± 2.14%
Host DNAs/RNAs	0.13 ± 0.06%	0.05 ± 0.01%	0.14 ± 0.05%
Reads mapped back contigs	78.47 ± 0.26%	66.85 ± 0.65%	54.43 ± 1.16%
eggNOG annotated reads	1,010,497 ± 32,603	23,590 ± 1,494	412,875 ± 30,166

1. QC, quality control.

Table 4.3 Relative abundance of differential microbial taxa between RFI groups in three breeds

	Taxon	H-RFI (Mean ± SEM)	L-RFI (Mean ± SEM)	*P* value[1]
Angus (ANG)				
Genus level	Unnamed RF19	0.49 ± 0.08%	1.31 ± 0.32%	0.045
Charolais (CHAR)				
Phylum level	*Firmicutes*	22.45 ± 2.14%	28.56 ± 1.82%	0.042
Phylum level	*Chloroflexi*	0.03 ± 0.01%	0.05 ± 0.01%	0.046
Genus level	Unclassified *Clostridiales*	3.05 ± 0.36%	4.38 ± 0.25%	0.008
Genus level	Unnamed *Ruminococcaceae*	1.20 ± 0.34%	2.15 ± 0.32%	0.040
Genus level	Unnamed S24-7	0.76 ± 0.20%	1.84 ± 0.44%	0.025
Genus level	*Succiniclasticum*	0.34 ± 0.15%	0.91 ± 0.10%	0.004
Genus level	Unnamed *Mogibacteriaceae*	0.38 ± 0.06%	0.58 ± 0.05%	0.023
Genus level	*Moryella*	0.29 ± 0.05%	0.56 ± 0.04%	0.002
Genus level	Unclassified *Clostridia*	0.20 ± 0.02%	0.32 ± 0.04%	0.012
Genus level	CF231	0.09 ± 0.02%	0.21 ± 0.03%	0.009
Genus level	Unclassified *Lachnospiraceae*	0.07 ± 0.01%	0.11 ± 0.01%	0.042
Genus level	p-75-a5	0.05 ± 0.01%	0.10 ± 0.02%	0.023
Genus level	Unclassified *Mogibacteriaceae*	0.05 ± 0.01%	0.08 ± 0.01%	0.049
Genus level	R4-45B	0.01 ± 0.01%	0.04 ± 0.01%	0.019
Genus level	*Blautia*	0.004 ± 0.001%	0.013 ± 0.003%	0.004
Genus level	*Adlercreutzia*	0.004 ± 0.001%	0.007 ± 0.001%	0.046
Genus level	Unclassified *Christensenellaceae*	0.004 ± 0.001%	0.007 ± 0.001%	0.043
Genus level	Unclassified *Anaerolineae*	0.004 ± 0.001%	0.005 ± 0.001%	0.018
Species level	Unclassified *Methanomassiliicoccaceae*	23.41 ± 1.31%	14.65 ± 2.96%	0.014
Species level	Unclassified *Methanobrevibacter*	5.48 ± 0.65%	8.24 ± 1.05%	0.033
Species level	Unclassified Group11	10.72 ± 0.86%	6.22 ± 1.69%	0.019
Species level	*Candidatus Methanomethylophilus alvus*	4.46 ± 0.37%	2.68 ± 0.75%	0.027
Crossbred (HYB)				
Genus level	Unclassified *Bacteroidales*	1.04 ± 0.13%	1.62 ± 0.09%	0.003
Genus level	Unclassified *Bacteroidetes*	0.86 ± 0.16%	1.73 ± 0.15%	0.001
Genus level	*Butyrivibrio*	0.97 ± 0.12%	1.58 ± 0.18%	0.009
Genus level	Unnamed *Victivallaceae*	0.77 ± 0.17%	1.70 ± 0.21%	0.006
Genus level	*Desulfovibrio*	0.21 ± 0.03%	0.43 ± 0.05%	0.002
Genus level	Unnamed *Mogibacteriaceae*	0.24 ± 0.02%	0.36 ± 0.05%	0.042
Genus level	Unclassified *Clostridia*	0.20 ± 0.02%	0.29 ± 0.02%	0.009
Genus level	Unnamed *Christensenellaceae*	0.07 ± 0.01%	0.25 ± 0.10%	0.036
Genus level	Unclassified *Paraprevotellaceae*	0.18 ± 0.03%	0.12 ± 0.01%	0.015
Genus level	*Shuttleworthia*	0.26 ± 0.09%	0.03 ± 0.01%	0.008
Genus level	Unnamed *Paraprevotellaceae*	0.19 ± 0.04%	0.10 ± 0.01%	0.045
Genus level	Unclassified *Aeromonadales*	0.15 ± 0.07%	0.01 ± 0.00%	0.027
Genus level	Unnamed *Spirochaetaceae*	0.04 ± 0.01%	0.10 ± 0.02%	0.015
Genus level	*Desulfobulbus*	0.10 ± 0.03%	0.03 ± 0.01%	0.009
Genus level	Unclassified *Verrucomicrobia*	0.04 ± 0.01%	0.07 ± 0.01%	0.045
Genus level	Unnamed R4-45B	0.02 ± 0.01%	0.07 ± 0.02%	0.008
Genus level	*Mitsuokella*	0.05 ± 0.01%	0.01 ± 0.00%	0.029
Genus level	L7A_E11	0.01 ± 0.00%	0.04 ± 0.01%	0.006
Genus level	*Roseburia*	0.04 ± 0.01%	0.01 ± 0.00%	0.020
Genus level	*Blautia*	0.01 ± 0.00%	0.04 ± 0.01%	0.035
Genus level	Unclassified *Lentisphaeria*	0.01 ± 0.00%	0.03 ± 0.01%	0.005
Genus level	Unclassified *Desulfovibrionaceae*	0.01 ± 0.00%	0.02 ± 0.01%	0.039
Species level	*Methanobrevibacter smithii*	2.04 ± 0.17%	2.95 ± 0.38%	0.040

1. *P* values were calculated between H- and L-RFI groups using *t* test within each breed, based on arcsine square root transformed relative abundances.

Table 4.4 Comparisons of alpha-diversity indices[1] between beef cattle with different RFI values

		Angus (ANG) (n = 16)	Charolais (CHAR) (n = 16) (mean ± SEM)	Crossbred (HYB) (n = 16)
Bacteria				
Chao1	L-RFI	317.03 ± 11.83	319.87 ± 12.95	381.88 ± 37.60
	H-RFI	337.64 ± 21.82	326.69 ± 16.70	358.74 ± 27.93
	P value[2]	NS	NS	NS
Shannon evenness	L-RFI	0.57 ± 0.01	0.59 ± 0.01	0.57 ± 0.01
	H-RFI	0.58 ± 0.01	0.55 ± 0.02	0.54 ± 0.01
	P value[2]	NS	0.03	NS
Simpson evenness	L-RFI	0.05 ± 0.00	0.05 ± 0.01	0.05 ± 0.00
	H-RFI	0.05 ± 0.00	0.04 ± 0.00	0.04 ± 0.00
	P value[2]	NS	NS	NS
Shannon index	L-RFI	3.17 ± 0.05	3.28 ± 0.06	3.21 ± 0.04
	H-RFI	3.18 ± 0.06	3.06 ± 0.10	3.01 ± 0.07
	P value[2]	NS	NS	0.04
Inverse Simpson	L-RFI	12.26 ± 0.96	13.89 ± 1.22	13.18 ± 0.74
	H-RFI	12.78 ± 0.97	10.96 ± 1.31	10.53 ± 1.14
	P value[2]	NS	NS	NS
Archaea				
Chao1	L-RFI	25.88 ± 1.03	28.06 ± 2.36	24.17 ± 0.95
	H-RFI	24.25 ± 1.22	24.65 ± 1.38	25.31 ± 1.32
	P value[2]	NS	NS	NS
Shannon evenness	L-RFI	0.64 ± 0.02	0.62 ± 0.02	0.66 ± 0.02
	H-RFI	0.64 ± 0.02	0.66 ± 0.01	0.65 ± 0.01
	P value[2]	NS	NS	NS
Simpson evenness	L-RFI	0.22 ± 0.02	0.20 ± 0.01	0.24 ± 0.02
	H-RFI	0.23 ± 0.02	0.26 ± 0.02	0.24 ± 0.01
	P value[2]	NS	0.03	NS
Shannon index	L-RFI	2.00 ± 0.07	1.95 ± 0.06	2.07 ± 0.07
	H-RFI	1.98 ± 0.07	2.07 ± 0.04	2.05 ± 0.03
	P value[2]	NS	NS	NA
Inverse Simpson	L-RFI	5.21 ± 0.51	4.74 ± 0.37	5.67 ± 0.46
	H-RFI	5.19 ± 0.50	5.93 ± 0.36	5.48 ± 0.14
	P value[2]	NS	0.03	NS

Note: [1]To make alpha-diversity indices comparable among samples, the number of sequences per sample was normalized to the lowest reads number for bacteria (n = 274,885) and archaea (n = 4,263), respectively.
[2]*P* values were obtained between H- and L-RFI groups within each breed using the Kruskal-Wallis rank sum test; NS, not significant with *P* value not less than 0.05.

Table 4.5 Abundance of differential functional categories between RFI groups in three beef cattle breeds

	Functional category	H-RFI (Mean±SEM; cpm[1])	L-RFI (Mean±SEM; cpm[1])	*P* value[2]
Angus (ANG)				
M-metatranscriptome	RNA processing and modification	49.84 ± 10.56	195.48 ± 103.74	0.021
Charolais (CHAR)				
metagenome	Cell cycle control, cell division, chromosome partitioning	12909.00 ± 347.45	11589.90 ± 301.67	0.008
metagenome	Secondary metabolites biosynthesis, transport and catabolism	12971.28 ± 700.02	11296.28 ± 508.41	0.033
T-metatranscirptome	Translation, ribosomal structure and biogenesis	96639.54 ± 3962.26	84353.42 ± 3284.80	0.026
T-metatranscirptome	Transcription	50431.27 ± 982.96	48084.54 ± 739.42	0.025
T-metatranscirptome	Coenzyme transport and metabolism	32933.44 ± 1089.03	35990.82 ± 1219.48	0.046
T-metatranscirptome	Chromatin structure and dynamics	874.42 ± 160.48	1409.80 ± 122.53	0.041
M-metatranscirptome	Translation, ribosomal structure and biogenesis	92501.13 ± 4666.25	72050.06 ± 3878.38	0.001
M-metatranscirptome	Coenzyme transport and metabolism	33326.68 ± 825.36	38128.97 ± 1781.97	0.014
Crossbred (HYB)				
metagenome	Intracellular trafficking, secretion, and vesicular transport	16275.15 ± 367.52	14238.87 ± 417.81	0.014
T-metatranscirptome	Cell motility	24372.19 ± 1793.06	33107.68 ± 5178.07	0.044
T-metatranscirptome	Nucleotide transport and metabolism	28338.08 ± 1255.51	24255.68 ± 1009.29	0.010
T-metatranscirptome	Cytoskeleton	4102.77 ± 1936.53	4132.47 ± 3014.61	0.036
M-metatranscirptome	Cell motility	24493.50 ± 1808.59	37334.38 ± 6269.31	0.013

1. cpm, counts per million reads.
2. *P* values were obtaining between H- and L-RFI steers using DESeq2 within each breed.

4.7 Figures

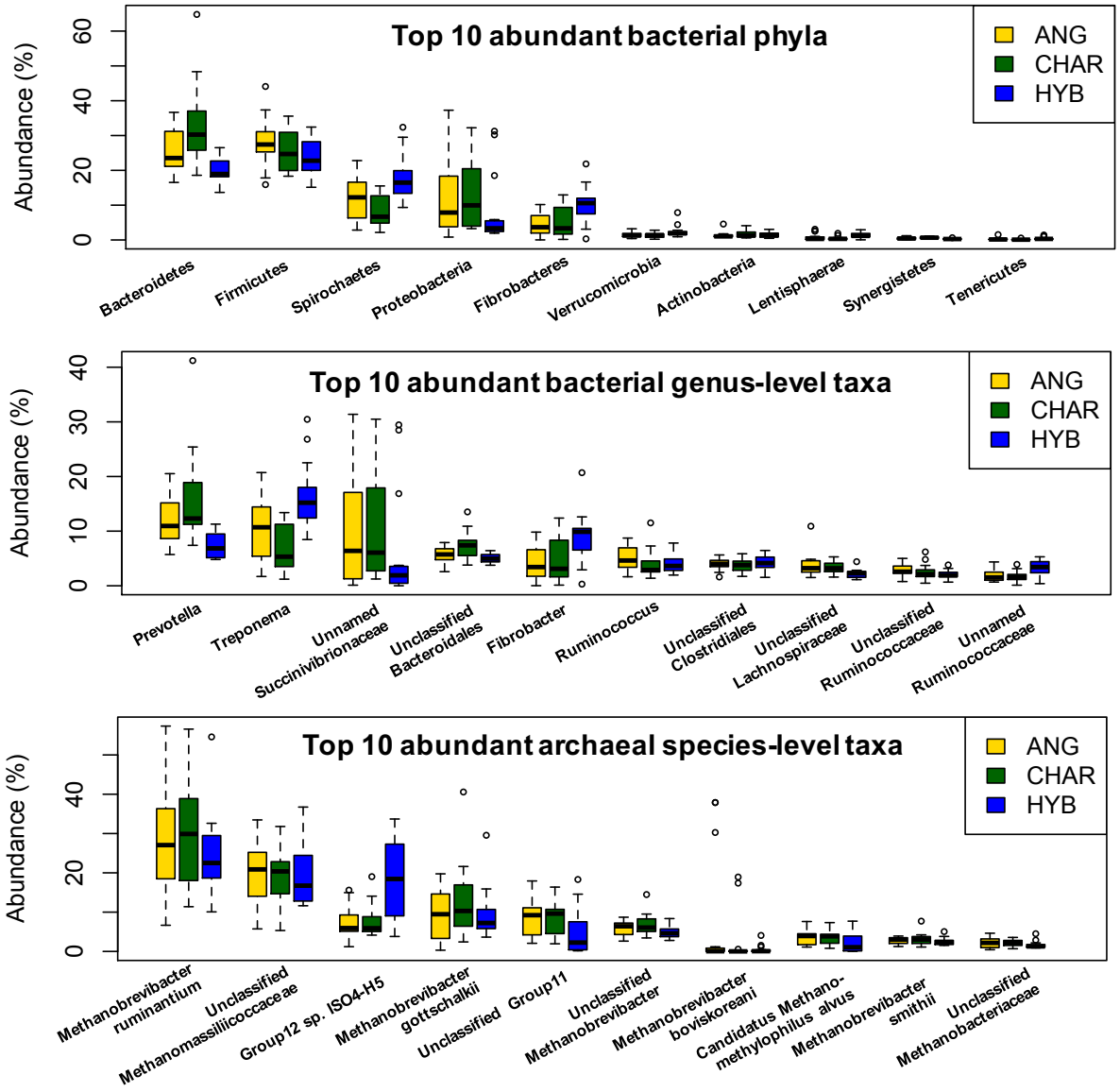


Figure 4.1 Relative abundance of the most abundant (top ten) rumen microbial taxa (at phylum and genus levels for bacteria and at the species level for archaea) among three beef cattle breeds. ANG = Angus, CHAR = Charolais, HYB = Kinsella composite hybrid.

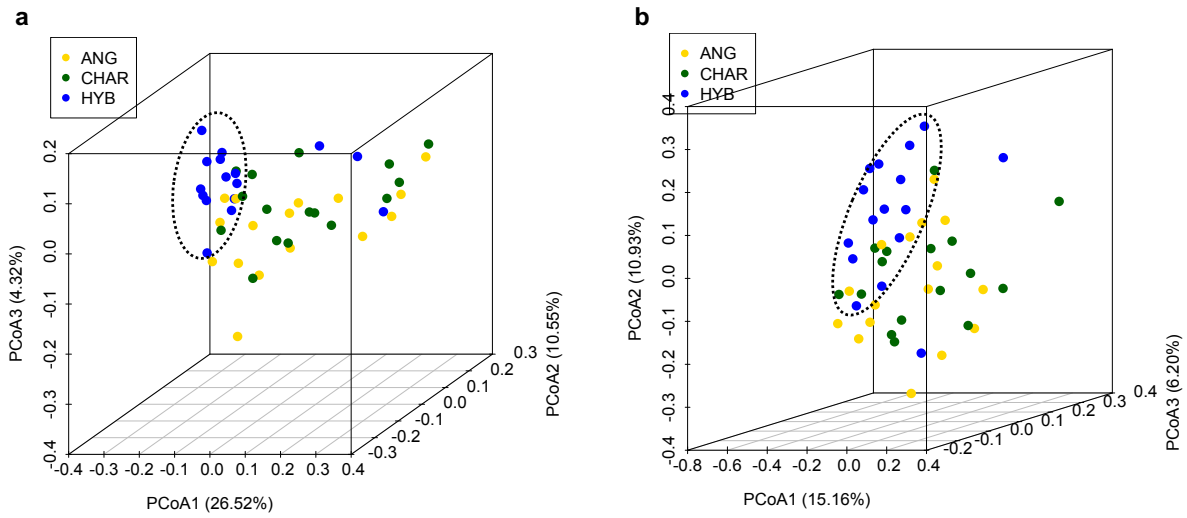


Figure 4.2 Rumen microbial compositional profiles of three beef cattle breeds visualized using the principal coordinate analysis (PCoA). The PCoA was conducted at the bacterial genus level and at the archaeal species level separately, based on Bray-Curtis dissimilarity matrices. The top three PCoAs were plotted for bacterial (**a**) and archaeal (**b**) communities.

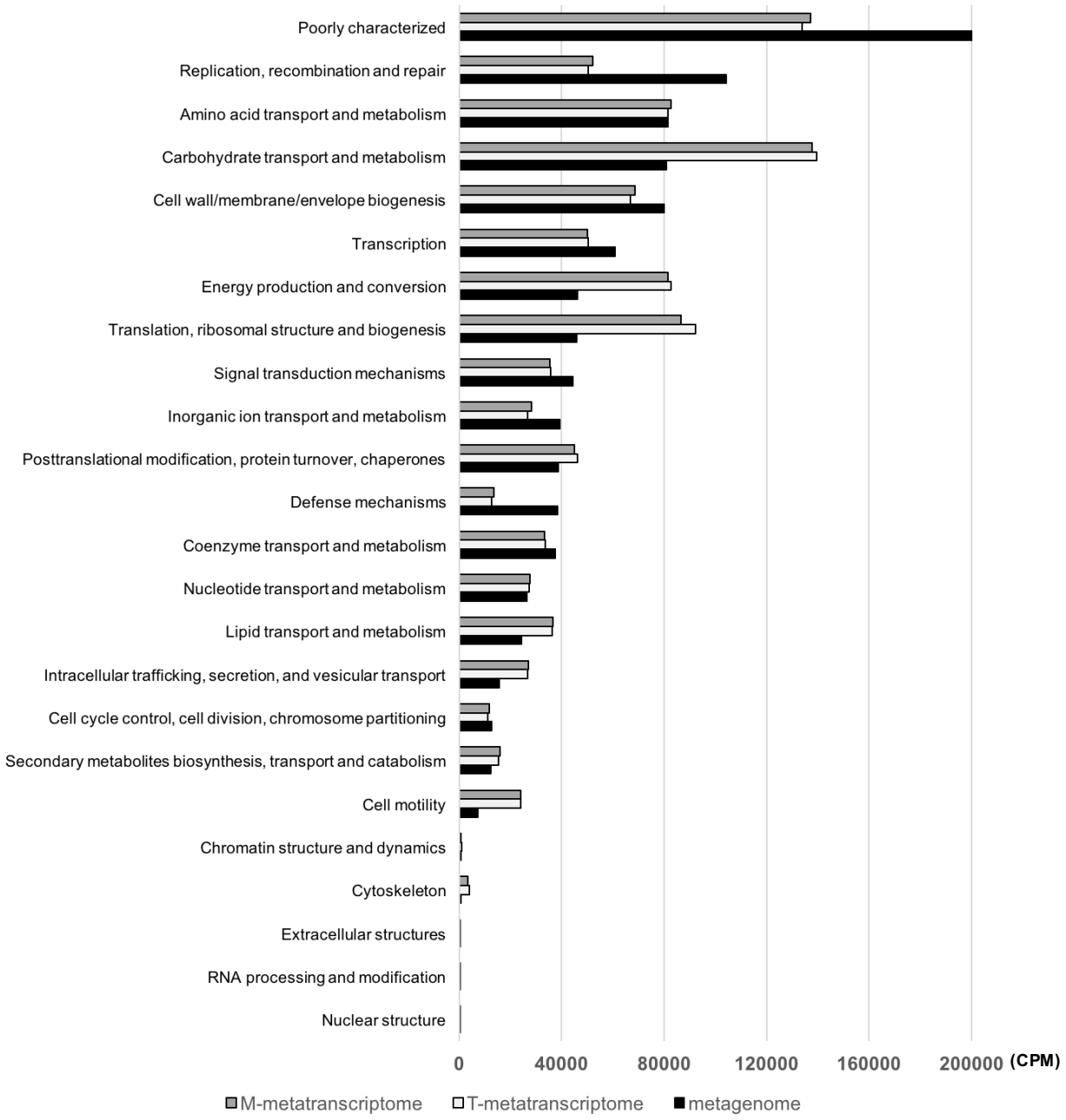


Figure 4.3 Abundance of observed eggNOG functional categories among metagenome, T-metatranscriptome, and M-metatranscriptome datasets. T- and M-metatranscriptome represents total-RNA-based and mRNA-enriched metatranscriptome, respectively.

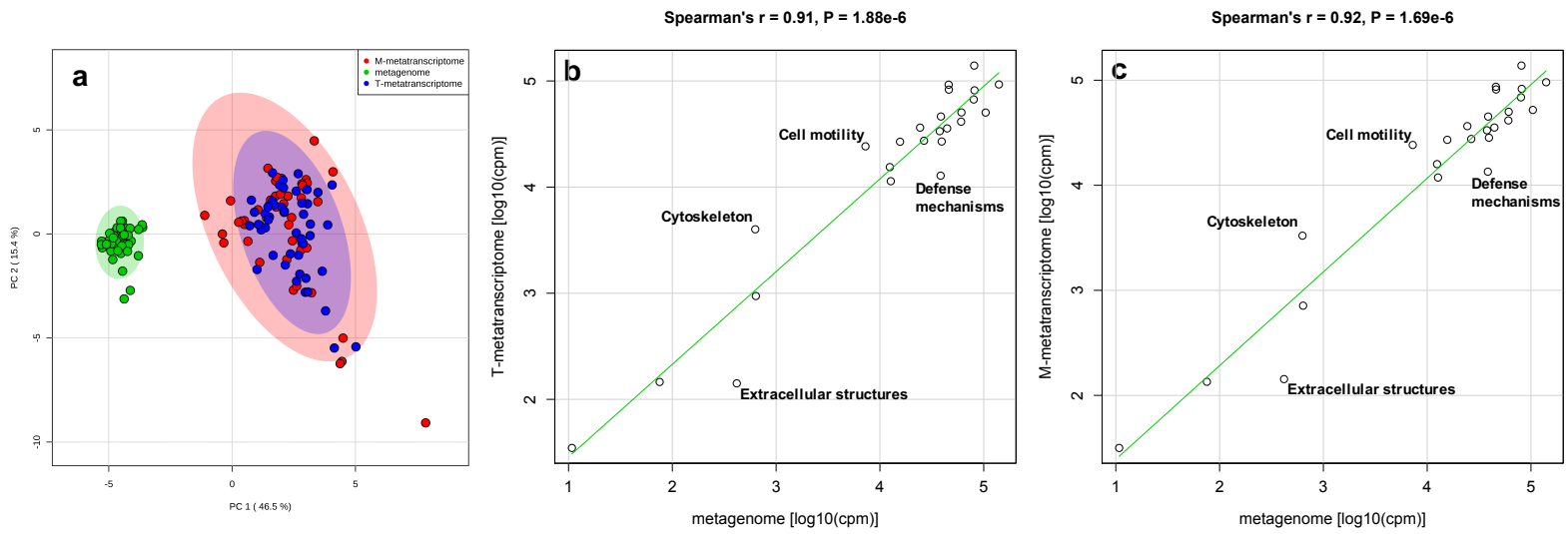


Figure 4.4 Distinguishable microbial functional profiles between rumen metagenomes and metatranscriptomes. (**a**) PCA for eggNOG functional categories, which was calculated based on auto-scaled abundances (cpm) of functional categories. (**b**) Correlation between metagenome and T-metatranscriptome. (**c**) Correlation between metagenome and M-metatranscriptome. Each scatterplot in **b** and **c** illustrates log10-transformed mean abundances (cpm) of each functional category at DNA and RNA levels.

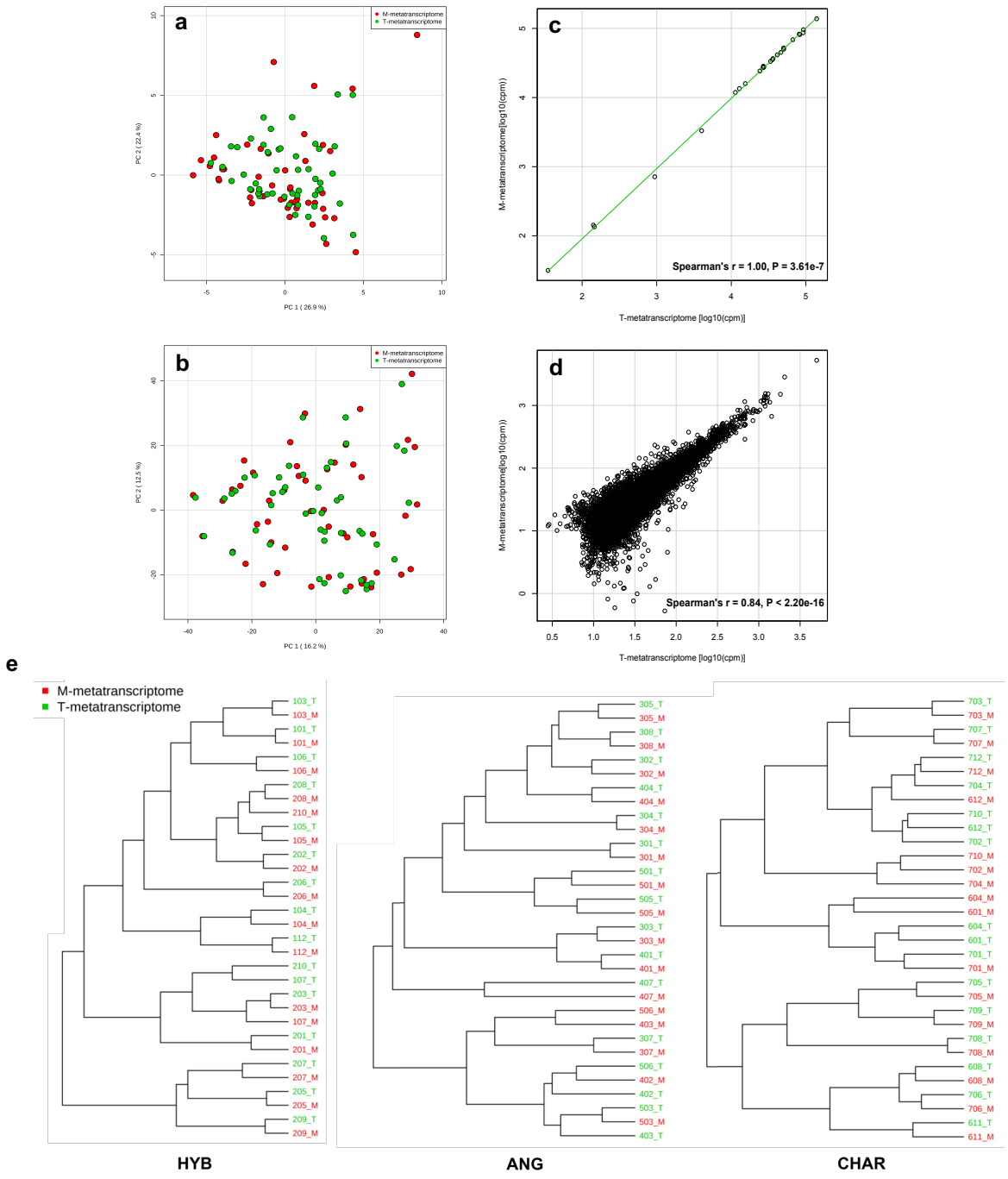


Figure 4.5 Microbial functional profiles of T- and M-metatranscriptomes. PCA for eggNOG functional categories (**a**) and expressed genes (**b**), which were performed based on auto-scaled abundances (cpm) of functional features. Correlations between rumen T- and M-metatranscriptomes were calculated using Spearman's rank correlation based on functional categories (**c**) and expressed genes (**d**). Each scatterplot in **c** and **d** illustrates log10-transformed mean abundances (cpm) of each functional category and each expressed gene. (**e**) Cluster analysis showing that between-method variations were lower than between-subject variations, which was conducted based on auto-scaled abundances (cpm) of functional categories using Euclidean as distance measure and Ward as clustering method.

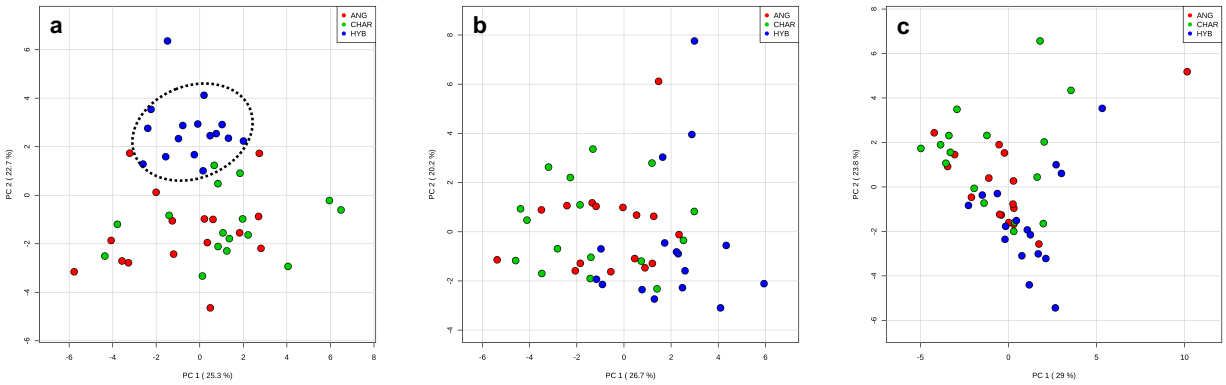


Figure 4.6 Microbial functional profiles of three beef cattle breeds. PCA for eggNOG functional categories from metagenomic (**a**), T-metatranscriptomic (**b**), and M-metatranscriptomic (**c**) datasets, which were performed based on auto-scaled abundances (cpm) of functional features.

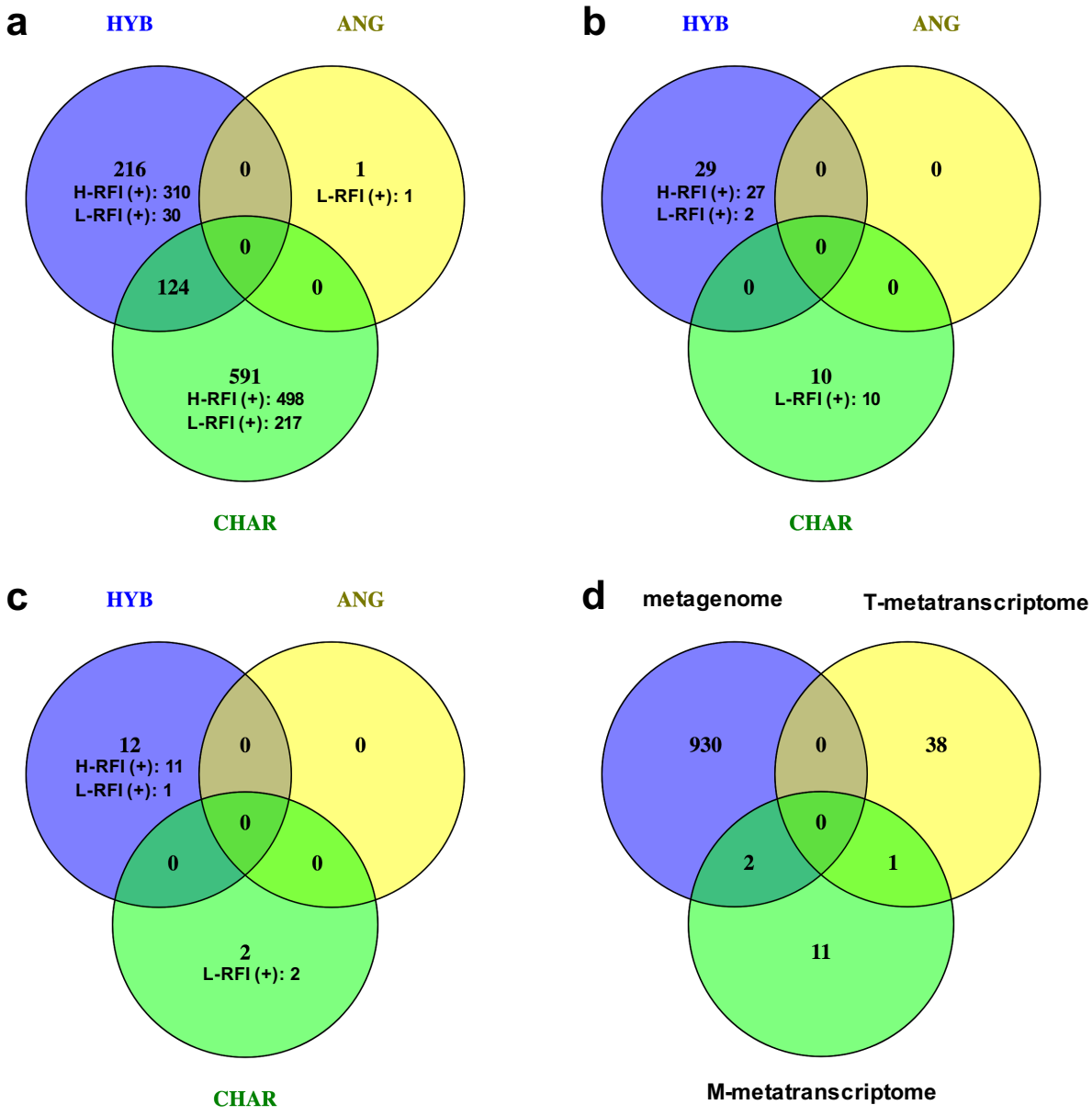


Figure 4.7 Identified differential genes/transcripts between H- and L-RFI groups from metagenome (**a**), T-metatranscriptome (**b**), and M-metatranscriptome (**c**) datasets, as well as differential genes/transcripts between RFI groups in all three datasets (**d**). H-RFI (+) and L-RFI (+) represents the number of genes/transcripts enriched in H-RFI and L-RFI animals, respectively.

Chapter 5. Unraveling genetic determinants for the rumen microbiota: insights for breeding more sustainable cattle

5.1 Introduction

The global human population is estimated to reach 9.15 billion by 2050, and is projected to result in an increased demand for livestock meat and milk products as a major protein source in the diet (Thornton, 2010). The majority of the world's milk and a high proportion of meat come from ruminants, mostly cattle, goats and sheep, highlighting the importance of ruminant animal production in the agriculture sector. Although cattle productivity has been improved through feeding cattle with cereal grains, such production systems will face the challenge of potential competitions for these inputs with humans (Eisler *et al.*, 2014). In addition, increasing beef production through raising more animals requires more arable land and water (Thornton, 2010), and it also poses significant environmental issues. Cattle produce enteric methane (CH_4) and nitrogenous compounds through urine and manure (Morgavi *et al.*, 2013), which contribute to greenhouse gases as well as environmental pollution. Thus, improving production efficiency is one of the priorities for mitigating these impacts whilst satisfying the increased demand for beef producers.

Ruminants have evolved to possess a diverse symbiotic microbiota in their rumen. The rumen microbiota, mainly consisting of bacteria, archaea, ciliated protozoa, fungi, and viruses (Firkins and Yu, 2015), can degrade complex plant polysaccharides and produce volatile fatty acids (VFAs), microbial proteins, and vitamins, which provide nutrients to meet the host's requirement for maintenance and growth. Using omics-based approaches, recent studies have suggested that differences in rumen microbiota were associated with cattle production and health traits, such as feed efficiency (Shabat *et al.*, 2016; Chapter 3 and 4), CH_4 yields (Wallace *et al.*,

2015), milk composition (Jami *et al.*, 2014), and ruminal acidosis (McCann *et al.*, 2016). Hence, the rumen microbiota is a potential target for manipulation to improve ruminant productivity and animal health, as well as to reduce CH_4 emissions.

It has been commonly accepted that diet plays the main role in shaping the gut microbiota (Spor *et al.*, 2011); however, more and more evidence from quantitative genetics, especially genome-wide association studies (GWAS) have revealed that host genetics is also important in determining the composition of gut microbiota in humans and mice. For example, 18 quantitative trait loci (QTLs) were found to be associated with the abundance of mouse gut microbial taxa (Benson *et al.*, 2010), and a follow-up study reported 42 QTLs for the abundance of 39 microbial taxa in a different mouse strain (Leamy *et al.*, 2014). Two studies conducted to estimate heritability of gut microbial taxonomic features in human found that the abundance of one-third of the identified operational taxonomic units (OTUs) were heritable with moderate or high heritability estimates (Goodrich *et al.*, 2014; Goodrich *et al.*, 2016). In addition, substantial associations between specific host genes and the gut microbiota were observed in the UK human population using GWAS. Examples included, the gene *SLIT3* (slit guidance ligand 3) with unclassified *Clostridiaceae*, *R3HDM1* (R3H domain containing 1) with *Bifidobacterium*, and *UHRF2* (ubiquitin like with PHD and ring finger domains 2) with weighted UniFrac distance (a measure of the similarity) (Goodrich *et al.*, 2016). Recently, 58 single nucleotide polymorphisms (SNPs) were also reported to be associated with the abundance of 33 gut microbial taxa in human (Turpin *et al.*, 2016).

In ruminants, early studies have indicated that the rumen microbiota could be influenced by host breed/species (Guan *et al.*, 2008; Henderson *et al.*, 2015; Hernandez-Sanabria *et al.*, 2013; Paz *et al.*, 2016; Roehe *et al.*, 2016). For instance differences in the composition of rumen microbiota were detected between Holstein and Jersey cows fed the same diet (Paz *et al.*, 2016).

However, lactation cycles and age effects were not considered and these factors do contribute to the variations of rumen microbiota (Jami *et al.*, 2013; Jewell *et al.*, 2015). In one of these studies researching the role of rumen microbiota in CH_4 emissions, host genetics was reported to affect the archaea:bacteria ratio in the rumen (Roehe *et al.*, 2016), but it was unclear whether host genetics affect the rumen microbial composition. In another recent survey of rumen microbiota of 742 rumen and foregut samples from 32 species or sub-species of ruminants and foregut fermenters across continents (Henderson *et al.*, 2015), the effects of diet, geographical regions, and genetic backgrounds of the host were confounded and could not be clearly separated. In two other studies investigating rumen microbiota from beef cattle with different feed efficiency (Guan *et al.*, 2008) and sire breeds (Hernandez-Sanabria *et al.*, 2013), although effects of breed (Guan *et al.*, 2008) and sire breed (Hernandez-Sanabria *et al.*, 2013) on the rumen microbiota were observed, rumen microbial communities were described using low resolution methods based on PCR-based fingerprinting method and qPCR. Currently, there is no in-depth understanding of how, and to what extent, host genetics can impact which microbial taxa because the host genetic effect on the rumen microbiota are poorly identified.

Recent studies highlight the individual variations of rumen microbiota in both beef (Chapter 3) and dairy cattle (Jami and Mizrahi, 2012) even when animals were fed the same diet and managed under the same environment, further suggesting an important role of host genetics in rumen microbial composition. In this study, we hypothesized that there are specific host genetic components (e.g., SNPs, genes, QTLs) contributing to the variations of microbial composition in the rumen, which could drive the "individualized" rumen microbiota. We assessed compositional profiles of the rumen microbiota, estimated the heritability, and performed GWAS for rumen microbial taxonomic features, through surveying a cohort of beef cattle (n = 712) raised under the same environment. With this extensive dataset and

comprehensive analyses, we aimed to determine the potential to breed beef cattle with a desirable rumen microbiota that contributes to high productivity and low environmental impact.

5.2 Material and Methods

5.2.1 Animal experiments and rumen sampling

A total of 712 beef cattle from three populations, including purebred Angus (ANG, n = 203) and Charolais (CHAR, n = 114) cattle, and Kinsella composite hybrid (HYB, n = 395), were raised under the same feedlot conditions at the Roy Berg Kinsella Research Ranch at the University of Alberta. The HYB population was bred from multiple beef breeds including Angus, Charolais, Galloway, Hereford, Holstein, Brown Swiss, and Simmental as described previously (Nkrumah *et al.*, 2007b). The experimental protocol was developed according to the guideline of the Canadian Council on Animal Care (Olfert *et al.*, 1993), and was approved by the Animal Care and Use Committee of the University of Alberta (protocol no. AUP00000882). Animals were fed with different diets according to their breed, sex, and growth stages (**Table 5.1**). Approximately 50 ml of rumen sample (including rumen fluid and feed particles) was collected from each animal using the oro-gastric tubing before feeding as previously described (Hernandez-Sanabria *et al.*, 2010), when the cattle were 293.0 ± 0.6 (mean ± SEM) days of age. Samples were immediately frozen using dry ice and then stored at -80 °C for further processing.

5.2.2 DNA extraction, high-throughput sequencing, and quantitative PCR (qPCR) analysis

Total DNA was isolated from each rumen sample using QIAGEN BioSprint 96 workstation (Valencia, California, United States) at Delta Genomics (Edmonton, AB, Canada). To assess the rumen bacterial and archaeal compositional profiles, the bacterial V1-V3 region and the archaeal V6-V8 region of 16S rRNA genes were amplified using primers as described

previously (Henderson *et al.*, 2015), i.e. for bacteria, the primers were Ba9F (5'-GAGTTTGATCMTGGCTCAG-3') and Ba515Rmod1 (5'-CCGCGGCKGCTGGCAC-3'); for archaea, the primers were Ar915aF (5'-AGGAATTGGCGGGGGAGCAC-3') and Ar1386R (5'-GCGGTGTGTGCAAGGAGC-3'). Regional amplicons were paired-end sequenced (2 × 300 bp) using the Illumina MiSeq PE300 at Génome Québec Innovation Centre (McGill University, Montréal, QC, Canada). All sequencing data are available for download from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession number PRJNA393057.

qPCR was performed to determine the abundance of rumen bacteria and archaea through enumerating their 16S rRNA gene copy numbers, using U2 primers for bacteria (forward: 5'-ACTCCTACGGGAGGCAG-3'; reverse: 5'-GACTACCAGGGTATCTAATCC-3') (Stevenson and Weimer, 2007) and uniMet1 primers for archaea (forward: 5'-CCGGAGATGGAACCTGAGAC-3'; reverse: 5'-CGGTCTTGCCCAGCTCTTATTC-3') (Zhou *et al.*, 2009). Standard curves were made using serial dilutions of plasmid DNA containing full length of 16S rRNA gene of *Butyrivibrio hungatei* (for U2 primers, using an initial concentration of 8.50×10^7 mol/µl) and partial 16S rRNA gene of *Methanobrevibacter sp.* strain AbM4 (for uniMet1 primers, using an initial concentration of 1.58×10^7 mol/µl). qPCR was conducted using SYBR Green chemistry (Fast SYBRH Green Master Mix; Applied Biosystems) in the StepOnePlus Real-Time PCR System (Applied Biosystems), and the 16S rRNA gene copy numbers per ml of rumen sample were calculated using the formula from a previous study (Zhou *et al.*, 2009).

5.2.3 Microbial composition analysis

Sequencing data were processed using the MacQIIME version 1.9.1. Briefly, paired-end

forward and reverse reads were joined, and then primers and homopolymer runs (maximum length: 8) of joined sequences were trimmed. Only sequences ≥ 400 bp in length, with the average quality score ≥ 25, and with ambiguous bases ≤ 6 were remained for downstream analysis. *De novo* chimera checking was performed using UCHIME (Edgar *et al.*, 2011) and operational taxonomic unit (OTU) picking was conducted using usearch (Edgar, 2010) to cluster similar sequences sharing ≥ 97% similarity. Representative sequences for bacterial and archaeal OTUs were assigned to the Greengenes 16S rRNA gene database (version gg_13_8) (McDonald *et al.*, 2012) and RIM-DB database (Seedorf *et al.*, 2014), respectively, using BLAST (Altschul *et al.*, 1990). Samples with < 500 bacterial sequences or samples with < 100 archaeal sequences were removed from the compositional analysis (Henderson *et al.*, 2015). To estimate Good's coverage and α-diversity indices (Chao1, Shannon index, and Simpson index), bacterial and archaeal sequences per sample were normalized to 2,000 and 500, respectively, using 100 subsampling iterations. These α-diversity indices were calculated at the genus level for bacterial communities, and at the species level for archaeal communities. β-diversity (Principal Coordinates Analysis [PCoA]) was calculated based on normalized sequence numbers (n = 2,000 for bacteria and n = 500 for archaea) using Bray-Curtis dissimilarity matrices. Samples with read number less than these cut-offs were not included in the analysis. The sequencing data were summarized at five taxonomic levels (from genus to phylum) for bacteria, and at six taxonomic levels (from species to phylum) for archaea. Only taxa with a relative abundance > 0.5% in at least one of samples were considered as detected taxa and included into the downstream analysis (Henderson *et al.*, 2015).

5.2.4 Co-occurrence network of rumen microbiota

Correlations between detected bacterial OTUs at the genus level and archaeal OTUs at the species level were inferred using the SparCC program (Friedman and Alm, 2012) implemented in mother (Schloss *et al.*, 2009), with default settings apart from "permutations = 10000". To avoid the potential bias on the co-occurrence calculations caused by different sequencing depths among samples, bacterial and archaeal OTU tables were subsampled to 2,000 and 500 reads for each sample, respectively, and samples with read number less than these cut-offs were removed from the downstream analysis. Bacterial and archaeal OTUs that were found in < 20% of animals in the population were also eliminated as previously suggested (Ramayo-Caldas *et al.*, 2016). The correlation patterns were further filtered to select only correlations with coefficient > 0.3 or < -0.3 and with *P* value < 0.001, which were then displayed using Cytoscape (Shannon *et al.*, 2003).

5.2.5 Genotyping

Genomic DNA was extracted from ear tissue of each animal, and genotyping was performed for all 712 beef cattle using the Illumina BovineSNP50 v2 Genotyping BeadChip containing 54,609 SNPs (San Diego, CA, United States) at Delta Genomics (Edmonton, AB, Canada). All 712 individuals were successfully genotyped with genotypes > 80%. Quality control for SNPs was then performed according to the following criteria: a) *P* value of Chi-square test of Hardy-Weinberg equilibrium > 10^{-6}; b) minor allele frequency (MAF) < 5%; c) genotyping call rate < 90%. Missing genotypes were imputed using the R package synbreed (Wimmer *et al.*, 2012). After filtering and imputation, 42,809 SNPs remained to construct the genomic relationship matrix (G) which was used in an animal model to estimate the heritability. In total 42,374 SNPs with known chromosomal position were used for GWAS (**Table 5.2**).

5.2.6 Heritability estimations

Only animals with completed rumen microbial profiles, phenotypic records and genotype information were included in this study. The relative abundance values of each microbial taxon were log10-transformed (Benson *et al.*, 2010). All values were plotted and possible outliers (out of mean ± 3SD) were removed, resulting in a total of n = 646 ~ 668 animals in the analyses for each microbial feature. To capture the additive genetic relationships among individuals, the genomic relationship matrix (G) was constructed based on the SNPs after quality control (n = 42,809) using the method previously developed (VanRaden, 2008) in the R package synbreed (Wimmer *et al.*, 2012). The heritability of each rumen microbial feature was estimated using the following animal model in ASReml (Gilmour *et al.*, 2014):

$$y_{ijklm} = \mu + b_i + s_j + d_k + g_l + a_m + e_{ijklm} \qquad [1]$$

Where y_{ijklm} is the microbial feature including log10-transformed abundance, alpha-diversity indices, and the top five bacterial/archaeal PCoAs from the Bray-Curtis matrices based PCoA as listed in **Table 5.4**; μ is the overall mean; b is the fixed breed effect with 3 classes (ANG, CHAR, and HYB); s is the fixed effect explaining differences between bull, heifer, and steer; d is the fixed effect of four different diets; g is the covariate representing the age effect at sampling, a is the random additive genetic effect following a distribution of $N(0, G\sigma_a^2)$, with the genomic relationship matrix G and the additive genetic variance σ_a^2; e is the random residual effect following $N(0, I\sigma_e^2)$, with identity matrix I and residual variance σ_e^2. The heritability (h^2) was defined as:

$$h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2) \qquad [2]$$

5.2.7 Genome-wide association studies (GWAS)

Firstly, microbial taxonomic features were adjusted for the fixed effects and covariate, including *breed*, *sex*, *diet*, and *age*. SNP positions were obtained using the SNPchiMp v.3 web-based tool (Nicolazzi *et al.*, 2015), and only SNPs with known positions (n = 42,374) were kept for the analysis. These SNPs were located on 30 *Bos taurus* chromosomes (29 autosomes [BTA] and the X chromosome; **Table 5.2**). GWAS were performed using rrBLUP (Endelman, 2011) in R package as the model below:

$$y^*_{ij} = \mu + a_i + m_j + e_{ij} \quad [3]$$

Where y^*_{ij} is the adjusted values of microbial taxonomic features; *a* and *e* is the random additive genetic effect and the random residual effect, respectively, with assumptions of distribution, variance and covariance structure as descripted above in model [1]; *m* is a fixed effect modeling the additive SNP effect. SNP genotypes were coded as -1/0/1 for genotype aa/Aa/AA. For each trait, *P*-values from testing the SNP effects were adjusted into genome-wide false discovery rates (FDRs) using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Associations with FDR < 0.1 were considered significant, and associations with 0.1 < FDR < 0.2 were regarded as suggestively significant.

5.3 Results

5.3.1 Survey of rumen microbiota using a large cohort of beef cattle

Rumen microbiota were surveyed using a cohort consisting of bulls (n = 71), heifers (n = 350) and steers (n = 291) that were born in 2014 and raised at the Roy Berg Kinsella Research Ranch at the University of Alberta. An average of 8,016 ± 96 (mean ± SE) and 1,862 ± 22

quality-filtered sequences were generated per animal for bacteria and for archaea, respectively, which were binned into a total of 67,791 bacterial and 168 archaeal OTUs (≥ 97% similarity). Good's coverages for both bacterial and archaeal communities were higher than 99% (**Table 5.3**). After classifying these OTUs into different taxonomic levels, 17 phylum-level taxa, 21 class-level taxa, 24 order-level taxa, 37 family-level taxa, and 64 genus-level taxa were detected for bacterial communities (with a cutoff for relative abundance > 0.5% in at least one of samples as suggested previously) (Henderson *et al.*, 2015). Meanwhile, one phylum-level taxa, three class-level taxa, three order-level taxa, three family-level taxa, ten genus-level taxa, and 14 species-level taxa were detected for archaeal communities (**Supplementary Table S5.1**). The dominant bacterial phyla were *Bacteroidetes* (44.05%), *Firmicutes* (36.42%), and *Proteobacteria* (4.61%), and each of the remaining 14 minor phyla accounted for < 1.00% of abundance. The most abundant archaeal taxa were *Methanobrevibacter gottschalkii* (85.09%) and *Methanobrevibacter ruminantium* (9.91%), followed by members of *Methanomassiliicoccaceae* (3.49%) (**Figure 5.1** and **Supplementary Table S5.1**). From those 64 bacterial genus-level taxa and 14 archaeal species-level taxa, *Prevotella*, unclassified *Ruminococcaceae*, unclassified *Clostridiales*, unclassified *Bacteroidales*, unclassified *Lachnospiraceae*, and *Methanobrevibacter gottschalkii* were found in more than 95% of the animals.

5.3.2 Factors driving segregation of rumen microbiota

General community structures (Principal Coordinates Analysis [PCoA] based on Bray-Curtis dissimilarity metrics), alpha-diversity indices (Chao1 for richness and Shannon for evenness), and abundance (16S rRNA gene copy numbers from qPCR) of rumen bacterial and archaeal communities were affected by breed, sex, and diet, while the age effect was only detected for the richness and abundance of bacteria communities (**Figure 5.2** and **5.3** and

Supplementary Table S5.1). From 197 detected bacterial and archaeal taxa, 50% (99), 91% (180), 92% (181), and 11% (22) of them were affected by breed, sex, diet, and age ($P < 0.05$), respectively (**Supplementary Table S5.1**).

Specific to the observed breed effect, both bacterial and archaeal profiles differed between Angus (ANG) and Charolais (CHAR) breeds of cattle, while those from the Kinsella composite hybrid (HYB) were overlapped with the two pure breeds (**Figure 5.2**). CHAR microbiota (bacterial and archaeal) were less diverse (with the lowest Chao1 and Shannon indices) than those of ANG and HYB (**Figure 5.3a-d**), while ANG microbiota had the highest richness (Chao1; **Figure 5.3a** and 5.**3b**). Meanwhile, a similar level of bacterial abundance was detected among the three breed populations ($P = 0.15$), with higher archaeal abundance for HYB compared with those in CHAR and ANG ($P = 2.7e-4$; **Figure 5.3e** and **5.3f**).

PCoA also displayed a sex effect on the bacterial communities, with no obvious sex effect on archaeal communities (**Figure 5.2**). In addition, comparison analysis of alpha-diversities revealed that the bull rumen microbiota had the lowest richness and evenness for archaeal communities and highest richness for bacterial communities (**Figure 5.3a-d**). Among the three genders, bulls had the highest archaeal but lowest bacterial abundance, while steers had lowest archaeal but highest bacterial abundance (**Figure 5.3e** and **5.3f**).

5.3.3 Microbial interactions detected in the cattle rumen

Bacterial abundance was correlated with archaeal abundance (correlation coefficient [ρ] = 0.26, $P = 3.64e-12$; Spearman's rank correlation; **Figure 5.4a**). Meanwhile, the abundance of *Firmicutes* was negatively correlated with that of *Bacteroidetes* ($\rho = -0.83$, $P = 2.20e-16$; **Figure 5.4b**). The most abundant archaeal species *Mbb. gottschalkii* was also negatively related to the second abundant archaeal species *Mbb. ruminantium* ($\rho = -0.75$, $P = 2.20e-16$; **Figure 5.4c**). In

addition, co-occurrence networks were observed for the bacterial communities but not for the archaeal communities (**Figure 5.5**), with 72 significant associations (52 positive and 20 negative) (correlation coefficient < -0.3 or > 0.3 and $P < 0.001$) being identified between bacterial taxa at the genus level. Four major modules comprised of correlated bacterial taxa were observed, centered by unclassified *Clostridiales*, unclassified *Succinivibrionaceae*, unclassified *Coriobacteriaceae*, and unclassified *Christensenellaceae,* respectively (**Figure 5.5b-e**).

5.3.4 Heritability estimates of rumen microbiota

The proportion of rumen microbial taxon at multiple taxonomic levels was treated as an individual trait as suggested previously (Turpin *et al.*, 2016), and its heritability (h^2) was estimated using an animal model based on the genomic relationship matrix (G matrix). In the present study, only microbial taxonomic features with a heritability estimate of $h^2 \geq 0.15$ were considered as being heritable. The results showed that animal additive genetic variations contributed to relative abundance of 58 (55 for bacteria and 3 for archaea) microbial taxa ($h^2 \geq 0.15$; **Figure 5.1** and **Supplementary Table S5.1**) belonging to various taxonomic levels. Among those 58 heritable bacterial taxa, 21 of them belonged to the phylum *Firmicutes,* including *Ruminococcus* ($h^2 = 0.16 \pm 0.08$; mean ± SE), unclassified *Clostridiales* ($h^2 = 0.25 \pm 0.09$), *Blautia* ($h^2 = 0.18 \pm 0.08$), etc. However, most members belonging to *Bacteroidetes*, such as *Prevotella*, unclassified S24-7, and unclassified *Bacteroidales*, were less affected by host genetics ($h^2 < 0.15$). For the three heritable archaeal taxa, the heritability estimate was 0.23 ± 0.08 for *Methanobacterium*, 0.18 ± 0.08 for *Mbb. ruminantium*, and 0.23 ± 0.08 for *Methanobacterium alkaliphilum*.

In addition, rumen bacterial diversity indices, including Shannon index ($h^2 = 0.23 \pm 0.09$) and Simpson index ($h^2 = 0.19 \pm 0.08$), were also heritable (**Table 5.4**). Meanwhile, moderate

heritability estimates (h^2 = 0.15 ~ 0.25) were obtained for PCoA2 (5.13% of variation) and PCoA5 (2.40% of variation) of bacterial communities, and for PCoA1 and PCoA2 (35.19% and 22.31% of variation, respectively) of archaeal communities (**Table 5.4**). Moreover, moderate heritability was observed for the bacterial abundance (h^2 = 0.16 ± 0.07) but not for the archaeal abundance (h^2 = 0.05 ± 0.06) (**Table 5.4**). The ratio between *Firmicutes* and *Bacteroidetes* (h^2 = 0.15 ± 0.07), and the ratio between *Mbb. gottschalkii* and *Mbb. ruminantium* (h^2 = 0.17 ± 0.08) were also moderately heritable traits (**Table 5.4**), similar to the four major microbial network modules (h^2 = 0.15 ~ 0.30) (**Figure 5.5b-e**).

5.3.5 SNPs identified for rumen microbial taxonomic features through GWAS

When downstream GWAS were performed for microbial taxonomic features with $h^2 \geq$ 0.1, 19 SNPs located on BTA (*Bos taurus* autosome) 1, 2, 3, 5, 7, 10, 12, 13, 16, 19, 26, and 27 were identified to be associated with microbial taxonomic features at the significance level of false discovery rate (FDR < 0.1) or at the suggestive significance level of 0.1 < FDR < 0.2. Specifically, these SNPs were associated with the abundance of six bacterial genus-level taxa (unclassified BS11, *Ruminococcus*, unclassified *Lachnospiraceae*, YRC22, unclassified *Mogibacteriaceae*, and unclassified *Victivallaceae*), three bacterial families (BS11, *Paraprevotellaceae*, and *Victivallaceae*), one bacterial order (*Victivallales*), two bacterial classes (*Spirochaetes* and *Lentisphaeria*), and two bacterial phyla (*Spirochaetes* and *Lentisphaerae*) (**Table 5.5** and **Figure 5.6**). No significant (or suggestively significant) association was observed for alpha-diversity indices, PCoAs, bacterial and archaeal abundance, and relative abundance of archaeal taxa.

The most significant associations were BS11 family and unclassified BS11 at the genus level with the SNP: rs110670001 on BTA10 (P = 1.43e-07, FDR = 0.006). In addition, four

adjacent SNPs (rs110410597, rs41604961, rs109122489, and rs110469969) located in the region of 28.10 ~ 28.18 Mbp on BTA13, which were in complete linkage disequilibrium (data not shown), tended to be associated with the phylum *Spirochaetes* and the class *Spirochaetes* (P = 2.45e-05 ~ 2.69e-05, FDR = 0.17 ~ 0.19). Moreover, two genus-level taxa including unclassified *Lachnospiraceae* and *Ruminococcus* tended to be associated with one SNP (rs109961459 on BTA13; P = 2.61e-06, FDR = 0.11) and four SNPs (rs43235157 on BTA1, rs110461771 on BTA2, rs41656119 on BTA7, and rs110071335 on BTA10; P = 3.88e-06 ~ 1.80e-05, FDR = 0.16 ~ 0.19), respectively (**Table 5.5** and **Figure 5.6**).

5.4 Discussion

The dietary effect on the rumen microbiota has been largely reported (Henderson *et al.*, 2015), because different diets provide various available substrates for microorganisms with different niches due to their nutritional characteristics. However, knowledge of breed and sex effects on the rumen microbiota are scarce. In the present study, three beef cattle populations showed distinguishable bacterial and archaeal characteristics, supporting previous reports on the effects of host genetic background on the rumen microbiota (Guan *et al.*, 2008; Henderson *et al.*, 2015; Hernandez-Sanabria *et al.*, 2013; Paz *et al.*, 2016; Roehe *et al.*, 2016). For example, the breed effect on the general structure of rumen bacterial communities was firstly observed in beef cattle based on the PCR-based fingerprinting method and a small sample size (n = 31) (Guan *et al.*, 2008), which was confirmed by the PCoA analysis of the current study based a much larger cohort of 712 beef cattle. Furthermore, through this extensive dataset and comprehensive analyses, our study also confirmed several breed-associated rumen microbial features that were sporadically reported previously, such as the abundance of total rumen methanogens (Hernandez-Sanabria *et al.*, 2013), relative abundance of several microbial taxa (e.g. *Fibrobacter*, unclassified

Veillonellaceae) (Henderson *et al.*, 2015), archaea:bacteria ratio (Roehe *et al.*, 2016), alpha- and beta-diversities (Paz *et al.*, 2016).

It is known that phenotypic variations exist between different breeds in both dairy and beef cattle. For example, methane emissions were significantly different between two beef cattle breeds (Aberdeen Angus vs. Limousin) (Roehe *et al.*, 2016), and milk yield and milk composition differed between Holstein and Jersey dairy cattle (Capper and Cady, 2012). However, phenotypic variations could not be fully explained by genetic differences between these breeds, and it has been suggested that the rumen microbiota potentially contribute towards these variations (Paz *et al.*, 2016). According to our results, the CHAR population had the least diverse microbiota compared to HYB and ANG cattle. It has been reported that a less diverse rumen microbiota was linked to higher feed efficiency and/or lower CH_4 emissions in both dairy (Shabat *et al.*, 2016) and beef cattle (Zhou *et al.*, 2009). Indeed, a previous study found that CHAR beef steers consumed ~5% less daily dry matter intake (DMI) and had 2.7% more average dairy gain (ADG) than ANG beef steers (Mao *et al.*, 2013). Therefore, the greater feed efficiency performance in CHAR may be partially explained by the less diverse rumen microbiota of CHAR population, which represents a potential genetic superiority of this breed.

Although sex has also been suggested to be one of factors affecting the composition of gut microbiota in humans and mice (Davenport *et al.*, 2015; Org *et al.*, 2016), our current study is the first to evaluate the sex effect on the rumen microbiota. We found that the microbiota observed in bulls was distinguishable from that of heifers and steers. A recent study reported that male castration eliminated the gut microbial differences between males and females, and the hormone (e.g. testosterone) treatment prevented the changes of males after gonadectomy (Yurkovetskiy *et al.*, 2013). This suggests that differences in sex hormones could be one of elements to explain the variations among different genders, because sex hormones affected bile

acid profiles (Yurkovetskiy *et al.*, 2013) and the shifts of bile acid consequently influence gut microbiota (Li and Chiang, 2015). Meanwhile, males and females may be exposed to different environmental microbes due to different diets and different activities (Davenport *et al.*, 2015), and thus it could also in part drive the different microbial profiles between sexes. However, such a gender effect on the rumen microbiota raises several questions, especially in beef cattle. Most of the genetic improvement for productivity was achieved through breeding beef sires and passing their desirable characteristics to their offspring steers. Our previous study has suggested the sire breed had an effect on the frequency of particular rumen microbial phylotypes in their offspring steers, but the gender factor was not considered (Hernandez-Sanabria *et al.*, 2013). In the current study, three genders were included for each breed and this factor has now been shown to affect both rumen microbial community structures and relative abundance of many taxa. However, future research on comparing microbiota from multiple generations of beef cattle with different genders is needed to determine to what extent rumen microbiota in bulls could be inherited by their offspring. Recent human studies also highlight the potential vertical transmission of gut microbiota, especially from mothers to infants (Asnicar *et al.*, 2017). Therefore, to what extent cows may impact the rumen microbiota also needs to be explored since heifers have different rumen microbiota to bulls. It is notable that as our study was conducted in "real" beef cattle operations, and thus cattle with different genders were fed with different diets to fulfill their different energy requirements. Therefore, the sex effect detected can be nested or confounded with the dietary effect, which should be considered as a limitation of this study.

Estimating heritability of rumen microbial members helped identify the host genetic effect compared with environmental factors that shape the observed variations in rumen microbiota. To estimate heritability, microbial taxonomic features were regarded as complex polygenic traits in this study, as previously suggested (Benson *et al.*, 2010). The predominant bacterial phylum,

most of the bacterial taxa (20 out of 22) belonging to *Bacteroidetes* only had low heritability estimates, suggesting that members belonging to this phylum are largely affected by environmental factors, such as diet. Indeed, many studies have reported diet as the major factor determining the abundance of *Bacteroidetes*, *Prevotella*, unclassified *B*acteroidales, and so on (Henderson *et al.*, 2015). The results are in line with studies on human gut, in which taxa belonging *Bacteroidetes* were not heritable and showed obvious shifts under diet interventions (David *et al.*, 2014; Goodrich *et al.*, 2014).

On the other hand, phylum *Firmicutes* (the second most abundant phylum) and many taxa belonging to this phylum (21 out of 52) had moderate heritability estimates, suggesting that the host genetic effect contributes to the observed variations in this phylum. This is also consistent with a recent study of human gut microbiota (Goodrich *et al.*, 2016). Members belonging to *Firmicutes* play very important roles in fiber digestion and carbohydrate metabolism. For example, as the most abundant family in *Firmicutes*, *Ruminococcaceae* had moderate heritability. This family is composed of both fibrolytic organisms and members involved in starch hydrolysis, which could produce acetate, formate, succinate, and so on (Klieve *et al.*, 2007; Russell and Rychlik, 2001). Unclassified *Clostridiales* has been reported to be affected by both host and diet (Henderson *et al.*, 2015), and the moderate heritability estimate obtained in this study further confirmed the host genetic effect on its abundance. Although a previous study indicated that unclassified *Clostridiales* may play a role in biohydrogenation (Huws *et al.*, 2011), the ecology and functions of phylotypes belonging this group are largely unknown because most of them are uncultivable. In addition, unclassified *Clostridiales* was the hub of one co-occurrence network module, so the host may also impact the other taxa that have interactions with unclassified *Clostridiales* in the rumen. Therefore, isolating and characterizing members of unclassified *Clostridiales* could be a future research direction which could help define their ecological niches

in the rumen and reveal mechanisms between their interactions with the host and other rumen microorganisms. Regardless, the observed different heritability estimates between members of *Bacteroidetes* and *Firmicutes* suggest that host effects are not equal on different rumen microbial phylotypes. Therefore, genetic selection and breeding may be applied to alter rumen microbial taxa with moderate heritability estimates, whilst it is unlikely to have any effects on those members driven by environmental factors.

Coevolution might be one of the mechanisms to explain different host genetic effects on different rumen microbial taxa. For example, we found that the abundance of *Ruminococcus* was influenced by host genetics in the current study. It has been reported that members of this genus display large diversity and particular host-association patterns in different mammalian species (La Reau *et al.*, 2016), supporting the suggestion that there are coevolutionary relationships between *Ruminococcus* and the host. In addition, as major butyrate producers (e.g. *Butyrivibrio*, *Clostridium*, etc.) (Russell and Rychlik, 2001), most members of *Lachnospiraceae* (9 out of 10) were not heritable in the rumen, whereas most members of this family were reported to be heritable in the human gut (Goodrich *et al.*, 2014). This inconsistency of heritability estimates of *Lachnospiraceae* members between ruminant and human further suggests there are coevolutionary relationships between host and the gut microbiota. Further scanning and analysis of genomic characters of these heritable rumen taxa, such as the outcomes of the Hungate 1000 project (http://www.rmgnetwork.org/hungate1000.html), will provide more information to explain how host and rumen microorganisms have coevolved at the genomic level, and provide a better understanding of how host genetics shape these microbial taxa.

Furthermore, the identification of associations between host SNP genotypes and rumen microorganisms through GWAS provides further knowledge on which genetic components contribute to the variations of rumen microbiota of beef cattle. For instance, the SNP:

rs110461771 (associated with the variations in the abundance of *Ruminococcus*), is located within the gene *RAPH*1 (Ras Association (RalGDS/AF-6) and Pleckstrin Homology Domains 1) on BTA2. The *RAPH*1 gene is involved in cell migration, which is the function that has been suggested to be associated with the nutrient absorption abilities of the rumen epithelia in beef steers (Kong *et al.*, 2016). Therefore, polymorphism of the *RAPH*1 gene may contribute to differences in the rumen epithelial absorption of nutrients such as VFAs. The variations in ruminal epithelial VFA absorption have been reported to be associated with differences in ruminal pH (Aschenbach *et al.*, 2011) and the shift in ruminal pH could influence the rumen microbiota (Hernandez *et al.*, 2014). Another SNP: rs29003226 (associated with the abundance of YRC22), is close to the *CDC*7 (cell division cycle 7) gene on BTA3. The *CDC*7 gene encodes the cell division cycle protein with kinase activity and might be involved in cell division of rumen epithelium. It has been reported that increased cell division could increase the proportion of epithelial cells, papillae length, and papillae number (Xiang *et al.*, 2016), and the variations of these rumen physical structures are expected to have a potential influence on the rumen microbiota (Roehe *et al.*, 2016). In addition, the SNP: rs41911152 (associated with various microbial groups) (**Table 5.5**), is located upstream of *MYH*3 (Myosin Heavy Chain 3) on BTA19. The *MYH3* gene plays a role in muscle contraction (Racca *et al.*, 2015), and thus it may relate to rumen contraction frequency by affecting the muscle action of the rumen wall. Rumen contraction frequency is associated with the passage rate of rumen digesta which has been suggested to also influence the microbiota (Roehe *et al.*, 2016). Furthermore, expressions of all three genes in the rumen epithelial wall were detected in HYB beef steers raised under the same environment in our previous study (Kong *et al.*, 2016). Overall, the above microbiota-associated SNPs suggest that the host genetics driven rumen physical features and gene expressions could drive the composition of rumen microbiota. Future follow-up studies to evaluate the associations

between these genes and regions (using higher density SNP markers and/or gene sequencing) and rumen epithelial structure and thickness, passage rate, ruminal pH, and rumen microbiota will provide more direct evidence to support our suggestions.

It is notable that several rumen microbiota-associated SNPs overlap with known quantitative trait loci (QTLs) for several phenotypic traits in cattle, such as feed efficiency. For example, SNPs on BTA1 (rs109763257 and rs43235157), BTA5 (rs41257422), and BTA13 (rs110410597, rs41604961, rs109122489, and rs110469969) are located within the QTLs for average daily gain (ADG) (de Oliveira *et al.*, 2014; Li *et al.*, 2002; Rolf *et al.*, 2012). Meanwhile, SNPs on BTA3 (rs29003226), BTA19 (rs41911152), and BTA26 (rs110728224 and rs110448978) overlap with QTLs for residual feed intake (RFI, one of feed efficiency measures) (Sherman *et al.*, 2009). Such an overlap suggests that these QTLs may have pleiotropic effects on both rumen microbiota and feed efficiency, which may partly explain the associations between rumen microorganisms and RFI (Shabat *et al.*, 2016; Chapters 3 and 4). For instance, a pervious study reported associations between the unclassified [*Mogibacteriaceae*] and feed efficiency (Myer *et al.*, 2015a), and the QTL for feed efficiency on BTA26 overlaps with the QTL for unclassified [*Mogibacteriaceae*] in our study. This region may harbor a gene that affects both unclassified [*Mogibacteriaceae*] and feed efficiency, or the QTL may contain several linked genes that individually or simultaneously influence these two traits. In addition, it is also possible that host QTLs impact feed efficiency through effects on the rumen microbial composition. Further studies are required to confirm these cause-and-effect relationships behind these pleiotropic effects between the rumen microbiota and feed efficiency.

It is worth mentioning that moderate heritability estimates were detected for several feed efficiency and/or CH_4 emission related rumen microbial features reported in other studies, including the *Bacteroidetes*:*Firmicutes* ratio (related to CH_4 emissions) (Martinez-Fernandez *et*

al., 2016), Shannon index for bacterial communities (related to feed efficiency) (Shabat *et al.*, 2016), the abundance of *Ruminococcus* (related to feed efficiency) (Myer *et al.*, 2015a), *Blautia* (related to feed efficiency) (Myer *et al.*, 2015b), *Mbb. ruminantium* (related to feed efficiency and CH_4 emissions) (Danielsson *et al.*, 2017; Shabat *et al.*, 2016), and so on. These results highlight the potential to manipulate these heritable microbial taxonomic features through genetic selection and breeding, which represents a useful strategy to improve feed efficiency and to reduce CH_4 emissions. After developing genetic or genomic selection tools to predict genetic merit of host animals for traits including rumen microbial taxonomic features, feed efficiency, CH_4 emissions, and other rumen microbiota-related traits (e.g., milk composition, ruminal acidosis, etc.), ruminant productivity could be improved through altering the rumen microbiota by genetic selection and breeding. In addition, to manipulate those environmentally determined phylotypes with low heritability estimates (such as members belonging to *Bacteroidetes*, and most of archaeal taxa), individual feeding schemes should be considered. Therefore, it is important to combine both genetics-based (selection and breeding) and management-based (individual feeding schemes) approaches to achieve optimal host-microbiota-diet interactions and thus enhance the productivity of beef cattle.

In conclusion, this study assessed the determinant factors for the rumen microbiota, estimated the heritability of rumen microbial taxonomic features, and identified genetic components associated with specific rumen microbial taxa using samples collected from a large cohort of beef cattle (n = 712) raised under the same environment. Rumen microbiota of these beef cattle are generally consistent with those typically described at various taxonomic levels (Henderson *et al.*, 2015; Kim and Yu, 2014). Multiple factors, including breed, sex, and diet were identified to drive the variations of rumen microbiota among animals. The findings on moderate heritability estimates for rumen microbial taxonomic features and the identified microbial taxa

associated SNPs from GWAS are the first to show direct evidence that rumen microbial colonization in beef cattle can be affected by host genetics including breed genetic differences and animal additive genetic effects. It further highlights a potential to manipulate the rumen microbiota through genetic selection and breeding to enhance the productivity of ruminants.

5.5 References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

Aschenbach JR, Penner GB, Stumpff F, Gabel G (2011). Ruminant Nutrition Symposium: Role of fermentation acid absorption in the regulation of ruminal pH. *J Anim Sci* **89:** 1092-1107.

Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3:** e1029.

Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F *et al* (2017). Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *mSystems* **2**.

Benjamini Y, 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:** 289-300.

Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J *et al* (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107:** 18933-18938.

Capper JL, Cady RA (2012). A comparison of the environmental impact of Jersey compared with Holstein milk for cheese production. *J Dairy Sci* **95:** 165-176.

Danielsson R, Dicksved J, Sun L, Gonda H, Muller B, Schnurer A *et al* (2017). Methane Production in Dairy Cows Correlates with Rumen Methanogenic and Bacterial Community Structure. *Front Microbiol* **8:** 226.

Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y (2015). Genome-Wide Association Studies of the Human Gut Microbiota. *PLoS One* **10:** e0140301.

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE *et al* (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505:** 559-563.

de Oliveira PS, Cesar AS, do Nascimento ML, Chaves AS, Tizioto PC, Tullio RR *et al* (2014). Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genet* **15:** 100.

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194-2200.

Eisler MC, Lee MR, Tarlton JF, Martin GB, Beddington J, Dungait JA *et al* (2014). Agriculture: Steps to sustainable livestock. *Nature* **507:** 32-34.

Endelman JB (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome* **4:** 250-255.

Firkins JL, 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:** 1450-1470.

Friedman J, Alm EJ (2012). Inferring correlation networks from genomic survey data. *PLoS Comput Biol* **8:** e1002687.

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R, Butler D *et al* (2014). *ASReml user guide. Release 4.1 structural specification*: VSN International Ltd, Hemel Hempstead, UK.

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R *et al* (2014). Human genetics shape the gut microbiome. *Cell* **159:** 789-799.

Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C *et al* (2016). Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe* **19:** 731-743.

Guan LL, Nkrumah JD, Basarab JA, Moore SS (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol Lett* **288:** 85-91.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH (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.

Hernandez J, Benedito JL, Abuelo A, Castillo C (2014). Ruminal acidosis in feedlot: from aetiology to prevention. *ScientificWorldJournal* **2014:** 702572.

Hernandez-Sanabria E, Guan LL, Goonewardene LA, Li M, Mujibi DF, Stothard P *et al* (2010). Correlation of particular bacterial PCR-denaturing gradient gel electrophoresis patterns with bovine ruminal fermentation parameters and feed efficiency traits. *Appl Environ Microbiol* **76:** 6338-6350.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Zhou M, Moore SS, Guan LL (2013). Influence of sire breed on the interplay among rumen microbial populations inhabiting the rumen liquid of the progeny in beef cattle. *PLoS One* **8:** e58461.

Huws SA, Kim EJ, Lee MR, Scott MB, Tweed JK, Pinloche E *et al* (2011). As yet uncultured bacteria phylogenetically classified as Prevotella, Lachnospiraceae incertae sedis and

unclassified Bacteroidales, Clostridiales and Ruminococcaceae may play a predominant role in ruminal biohydrogenation. *Environ Microbiol* **13:** 1500-1512.

Jami E, Mizrahi I (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS One* **7:** e33306.

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

Jami E, White BA, Mizrahi I (2014). Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *PLoS One* **9:** e85423.

Jewell KA, McCormick CA, Odt CL, Weimer PJ, Suen G (2015). Ruminal Bacterial Community Composition in Dairy Cows Is Dynamic over the Course of Two Lactations and Correlates with Feed Efficiency. *Appl Environ Microbiol* **81:** 4697-4710.

Kim M, Yu Z (2014). Variations in 16S rRNA-based microbiome profiling between pyrosequencing runs and between pyrosequencing facilities. *J Microbiol* **52:** 355-365.

Klieve AV, O'Leary MN, McMillen L, Ouwerkerk D (2007). Ruminococcus bromii, identification and isolation as a dominant community member in the rumen of cattle fed a barley diet. *J Appl Microbiol* **103:** 2065-2073.

Kong RS, Liang G, Chen Y, Stothard P, Guan le L (2016). Transcriptome profiling of the rumen epithelium of beef cattle differing in residual feed intake. *BMC Genomics* **17:** 592.

La Reau AJ, Meier-Kolthoff JP, Suen G (2016). Sequence-based analysis of the genus Ruminococcus resolves its phylogeny and reveals strong host association. *Microb Genom* **2:** e000099.

Leamy LJ, Kelly SA, Nietfeldt J, Legge RM, Ma F, Hua K *et al* (2014). Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biol* **15:** 552.

Li C, Basarab J, Snelling WM, Benkel B, Murdoch B, Moore SS (2002). The identification of common haplotypes on bovine chromosome 5 within commercial lines of Bos taurus and their associations with growth traits. *J Anim Sci* **80:** 1187-1194.

Li F, Guan LL (2017). Metatranscriptomic Profiling Reveals Linkages between the Active Rumen Microbiome and Feed Efficiency in Beef Cattle. *Appl Environ Microbiol* **83**.

Li T, Chiang JY (2015). Bile acids as metabolic regulators. *Curr Opin Gastroenterol* **31:** 159-165.

Mao F, Chen L, Vinsky M, Okine E, Wang Z, Basarab J *et al* (2013). Phenotypic and genetic relationships of feed efficiency with growth performance, ultrasound, and carcass merit traits in Angus and Charolais steers. *J Anim Sci* **91:** 2067-2076.

Martinez-Fernandez G, Denman SE, Yang C, Cheung J, Mitsumori M, McSweeney CS (2016). Methane Inhibition Alters the Microbial Community, Hydrogen Flow, and Fermentation Response in the Rumen of Cattle. *Front Microbiol* **7:** 1122.

McCann JC, Luan S, Cardoso FC, Derakhshani H, Khafipour E, Loor JJ (2016). Induction of Subacute Ruminal Acidosis Affects the Ruminal Microbiome and Epithelium. *Front Microbiol* **7:** 701.

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A *et al* (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *Isme j* **6:** 610-618.

Morgavi DP, Kelly WJ, Janssen PH, Attwood GT (2013). Rumen microbial (meta)genomics and its application to ruminant production. *Animal* **7 Suppl 1:** 184-201.

Myer PR, Smith TP, Wells JE, Kuehn LA, Freetly HC (2015a). Rumen microbiome from steers differing in feed efficiency. *PLoS One* **10:** e0129174.

Myer PR, Wells JE, Smith TP, Kuehn LA, Freetly HC (2015b). Microbial community profiles of the colon from steers differing in feed efficiency. *Springerplus* **4:** 454.

Nicolazzi EL, Caprera A, Nazzicari N, Cozzi P, Strozzi F, Lawley C *et al* (2015). SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics* **16:** 283.

Nkrumah JD, Crews DH, Jr., Basarab JA, Price MA, Okine EK, Wang Z *et al* (2007). Genetic and phenotypic relationships of feeding behavior and temperament with performance, feed efficiency, ultrasound, and carcass merit of beef cattle. *J Anim Sci* **85:** 2382-2390.

Olfert ED, Cross BM, McWilliams AA (1993). Guide to the care and use of experimental steers. *Canadian Council on Animal Care, Ottawa, ON, Canada*.

Org E, Mehrabian M, Parks BW, Shipkova P, Liu X, Drake TA *et al* (2016). Sex differences and hormonal effects on gut microbiota composition in mice. *Gut Microbes* **7:** 313-322.

Paz HA, Anderson CL, Muller MJ, Kononoff PJ, Fernando SC (2016). Rumen Bacterial Community Composition in Holstein and Jersey Cows Is Different under Same Dietary Condition and Is Not Affected by Sampling Method. *Front Microbiol* **7:** 1206.

Racca AW, Beck AE, McMillin MJ, Korte FS, Bamshad MJ, Regnier M (2015). The embryonic myosin R672C mutation that underlies Freeman-Sheldon syndrome impairs cross-bridge detachment and cycling in adult skeletal muscle. *Hum Mol Genet* **24:** 3348-3358.

Ramayo-Caldas Y, Mach N, Lepage P, Levenez F, Denis C, Lemonnier G *et al* (2016). Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *Isme j* **10:** 2973-2977.

Roehe R, Dewhurst RJ, Duthie CA, Rooke JA, McKain N, Ross DW *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 Genet* **12:** e1005846.

Rolf MM, Taylor JF, Schnabel RD, McKay SD, McClure MC, Northcutt SL *et al* (2012). Genome-wide association analysis for feed efficiency in Angus cattle. *Anim Genet* **43:** 367-374.

Russell JB, Rychlik JL (2001). Factors that alter rumen microbial ecology. *Science* **292:** 1119-1122.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537-7541.

Seedorf H, Kittelmann S, Henderson G, Janssen PH (2014). RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. *PeerJ* **2:** e494.

Shabat SK, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Berg Miller ME *et al* (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *Isme j*.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498-2504.

Sherman EL, Nkrumah JD, Li C, Bartusiak R, Murdoch B, Moore SS (2009). Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. *J Anim Sci* **87:** 37-45.

Spor A, Koren O, Ley R (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* **9:** 279-290.

Stevenson DM, Weimer PJ (2007). Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl Microbiol Biotechnol* **75:** 165-174.

Thornton PK (2010). Livestock production: recent trends, future prospects. *Philos Trans R Soc Lond B Biol Sci* **365:** 2853-2867.

Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI *et al* (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat Genet* **48:** 1413-1417.

VanRaden PM (2008). Efficient methods to compute genomic predictions. *J Dairy Sci* **91:** 4414-4423.

Wallace RJ, Rooke JA, McKain N, Duthie CA, Hyslop JJ, Ross DW *et al* (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* **16:** 839.

Wimmer V, Albrecht T, Auinger HJ, Schon CC (2012). synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* **28:** 2086-2087.

Xiang R, McNally J, Rowe S, Jonker A, Pinares-Patino CS, Oddy VH *et al* (2016). Gene network analysis identifies rumen epithelial cell proliferation, differentiation and metabolic pathways perturbed by diet and correlated with methane production. *Sci Rep* **6:** 39022.

Yurkovetskiy L, Burrows M, Khan AA, Graham L, Volchkov P, Becker L *et al* (2013). Gender bias in autoimmunity is influenced by microbiota. *Immunity* **39:** 400-412.

Zhou M, Hernandez-Sanabria E, Le LG (2009). Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Appl Environ Microbiol* **75:** 6524-6533.

5.6 Tables

Table 5.1 Diet information for animal experiments

Sex	Breed	Diet	Diet composition
Steer	HYB	Diet A	95% straight oats and free choice hay, 5% rumensin pellet
	CHAR	Diet A	
	ANG	Diet B	Straight hay
Bull	HYB	Diet C	80% Silage, 15% whole oats, 5% rumensin pellet
	CHAR	Diet C	
	ANG	Diet C	
Heifer	HYB	Diet D	55% Silage, 40% whole oats, 5% rumensin pellet
	ANG	Diet D	
	CHAR	Diet A	95% straight oats and free choice hay, 5% rumensin pellet

Abbreviations: HYB = Kinsella composite hybrid, CHAR = Charolais, and ANG = Angus.

Table 5.2 Single nucleotide polymorphisms (SNPs) information

BTA	SNP[1] (n)	SNP interval (Mean, bp)	BTA	SNP[1]	SNP interval (Mean, bp)
1	2739	57741	16	1348	60253
2	2265	60363	17	1333	56221
3	2049	59151	18	1087	60222
4	2045	58885	19	1145	55519
5	1744	69466	20	1296	55146
6	2104	56607	21	1141	62367
7	1850	60766	22	1050	58357
8	1958	57695	23	889	58812
9	1700	62074	24	1058	58753
10	1788	58295	25	803	53258
11	1810	59224	26	899	56741
12	1391	65428	27	781	58015
13	1461	57440	28	812	56946
14	1472	56528	29	882	58005
15	1378	61325	X	96	1150102

#SNPs before quality control filter: 54609
#SNPs not following Hardy-Weinberg equilibrium: 1802
#SNPs with minor allele frequency (MAF) < 5% and genotyping call rate < 90%: 9998
#Available SNPs for genomic relationship matrix construction: 42809
#SNPs without position information: 435
#Available SNPs for genome wide association study: 42374

1. These numbers are the SNPs involved into the genome wide association study.

Table 5.3 Alpha-diversity indices[1] of beef cattle population

	Bacterial community[2]			**Archaeal community[2]**		
	Mean	SEM (n=658)	CV (%)	Mean	SEM (n=660)	CV (%)
Chao1	60.88	0.48	20.06	6.48	0.09	37.26
Shannon	3.00	0.02	15.60	0.63	0.02	62.72
Simpson	0.78	0.00	12.63	0.21	0.01	73.40
Good's coverage	99.33%	0.00	0.15	99.72%	0.00	0.14

1. To estimate these α-diversity indices, the number of bacterial and archaeal sequences per sample were normalized to 2,000 and 500, respectively, using 100 subsampling iterations.
2. These α-diversity indices were calculated at the genus level for bacterial communities, and at the species level for archaeal communities.

Table 5.4 Heritability estimates of rumen microbial abundance, diversity indices[1], and ratios between dominant microbial groups

Rumen microbial taxonomic features	Heritability ($h^2 \pm SE$)
Bacteria	
16S rRNA gene copy number (log10)	0.16 ± 0.07
Chao1 index	0.09 ± 0.07
Shannon index	0.23 ± 0.09
Simpson index	0.19 ± 0.08
PCoA1 (6.88% variation)	0.12 ± 0.07
PCoA2 (5.13% variation)	0.25 ± 0.09
PCoA3 (3.33% variation)	0.08 ± 0.06
PCoA4 (2.75% variation)	0.00 ± 0.00
PCoA5 (2.40% variation)	0.15 ± 0.09
Archaea	
16S rRNA gene copy number (log10)	0.05 ± 0.06
Chao1 index	0.00 ± 0.05
Shannon index	0.04 ± 0.06
Simpson index	0.05 ± 0.06
PCoA1 (35.19% variation)	0.17 ± 0.09
PCoA2 (22.31% variation)	0.17 ± 0.08
PCoA3 (6.18% variation)	0.05 ± 0.06
PCoA4 (4.58% variation)	0.00 ± 0.00
PCoA5 (2.76% variation)	0.06 ± 0.06
Ratio[2]	
Archaea: Bacteria	0.04 ± 0.06
Firmicutes: *Bacteroidetes*	0.15 ± 0.07
Mbb. gottschalkii: *Mbb. ruminantium*	0.17 ± 0.08

[1]To estimate these α- and β- diversity indices, the number of bacterial and archaeal sequences per sample were normalized to 2,000 and 500, respectively. α-diversity indices were calculated at the genus level for bacterial communities, and at the species level for archaeal communities. Principal Coordinates Analysis (PCoA) was conducted using Bray-Curtis dissimilarity matrices.
[2]Abundance from qPCR and relative abundance were both log10-transformed before we calculated these ratios.

Table 5.5 Identified bovine SNPs that associated with rumen microbial taxa

SNP	Position	Alleles	Gene	Consequence	Associated Taxon	FDR[3]	P
rs109763257	1:155345571	C/T	NC region[1]	NA[2]	*Spirochaetes* (Phylum)	0.173	1.20e-05
					Spirochaetes (Class)	0.190	9.43e-06
rs43235157	1:156294225	A/G	*TBC1D5*	Intron variant	*Ruminococcus* (Genus)	0.191	1.33e-05
rs110461771	2:92080445	C/T	*RAPH1*	Intron variant	*Ruminococcus* (Genus)	0.164	3.88e-06
rs29003226	3:51976646	C/G	NC region[1]	NA[2]	YRC22 (Genus)	0.107	2.53e-06
rs41257422	5:6266261	A/G	NC region[1]	NA[2]	YRC22 (Genus)	0.155	7.33e-06
rs41656119	7:83551608	A/G	NC region[1]	NA[2]	*Ruminococcus* (Genus)	0.191	1.80e-05
rs110670001	10:10930797	C/T	NC region[1]	NA[2]	BS11 (Family)	0.006	1.43e-07
					Unclassified BS11 (Genus)	0.006	1.43e-07
rs110071335	10:81981544	A/C	*SMOC1*	Intron variant	*Ruminococcus* (Genus)	0.191	1.46e-05
rs109402398	12:37678844	C/T	NC region[1]	NA[2]	*Paraprevotellaceae* (Family)	0.105	4.95e-06
rs110410597	13:28095457	C/T	*OPTN*	Intron variant	*Spirochaetes* (Phylum)	0.173	2.45e-05
					Spirochaetes (Class)	0.190	2.69e-05
rs41604961	13:28115879	C/T	*OPTN*	Intron variant	*Spirochaetes* (Phylum)	0.173	2.45e-05
					Spirochaetes (Class)	0.190	2.69e-05
rs109122489	13:28149879	C/T	*MCM10*	Intron variant	*Spirochaetes* (Phylum)	0.173	2.45e-05
					Spirochaetes (Class)	0.190	2.69e-05
rs110469969	13:28183389	C/T	*UCMA*	Intron variant	*Spirochaetes* (Phylum)	0.173	2.45e-05
					Spirochaetes (Class)	0.190	2.69e-05
rs109961459	13:24202640	A/G	NC region[1]	NA[2]	Unclassified *Lachnospiraceae* (Genus)	0.111	2.61e-06
rs41627213	16:78415671	C/T	*DENND1B*	Intron variant	*Paraprevotellaceae* (Family)	0.070	1.65e-06
rs41911152	19:30220186	C/T	NC region[1]	NA[2]	*Lentisphaerae* (Phylum)	0.070	1.64e-06
					Lentisphaeria (Class)	0.070	1.64e-06
					Victivallales (Order)	0.034	8.05e-07
					Victivallaceae (Family)	0.038	8.92e-07
					Unclassified *Victivallaceae* (Genus)	0.038	8.92e-07
rs110728224	26:32497450	A/G	NC region[1]	NA[2]	*Spirochaetes* (Phylum)	0.173	4.73e-06
					Spirochaetes (Class)	0.140	3.31e-06
rs110448978	26:37871121	C/T	*KCNK18*	Downstream variant	Unclassified *Mogibacteriaceae* (Genus)	0.187	4.40e-06
rs42620822	27:42776720	A/G	NC region[1]	NA[2]	*Spirochaetes* (Class)	0.196	3.24e-05

[1]NC region = non-coding region.
[2]NA = not available.
[3]For each microbial taxonomic feature, *P*-value was adjusted into genome-wide false discovery rates (FDRs) using the Benjamini-Hochberg method. Associations with FDR < 0.1 were considered significant, and associations with 0.1 < FDR < 0.2 were regarded as suggestively significant.

5.7 Figures

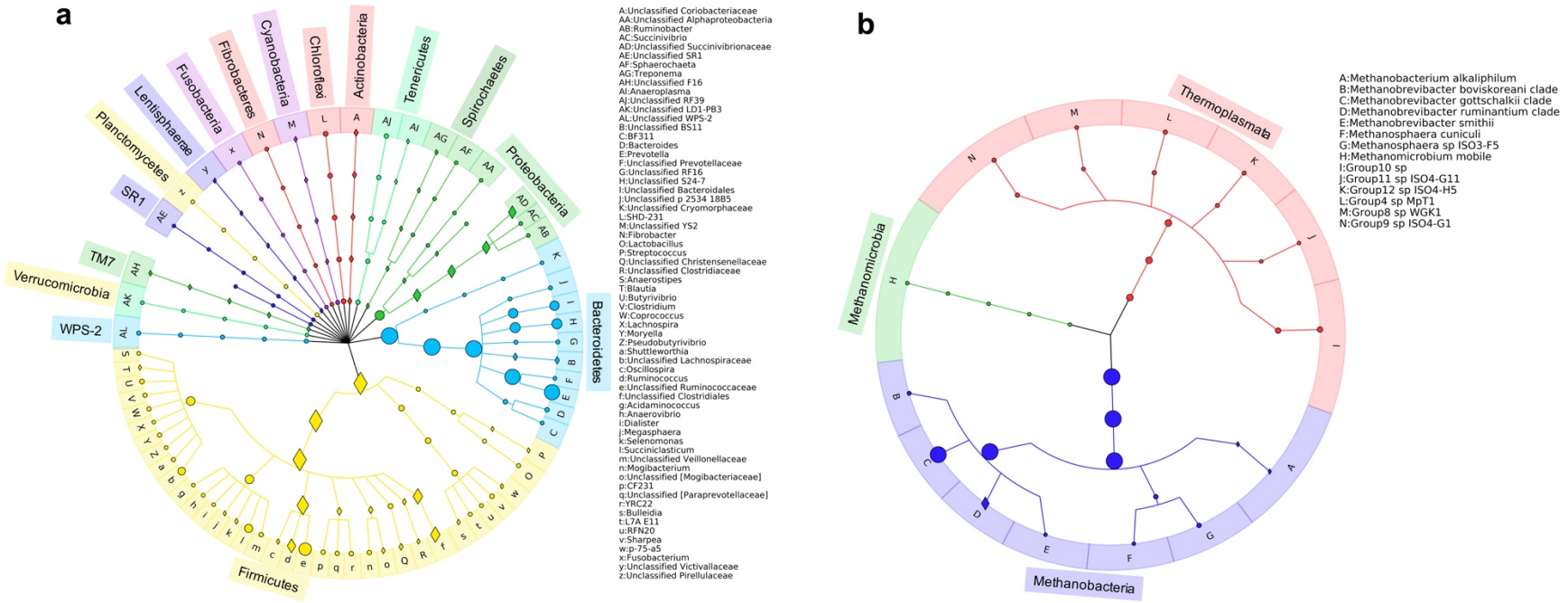


Figure 5.1 Composition of rumen microbiota in beef cattle. Bacterial community composition was summarized at genus, family, order, class, and phylum levels (**a**), and archaeal community composition was summarized at species, genus, family, order, and class levels (**b**). Heritable taxa (heritability estimate [h^2] $\geq$ 0.15) were indicated using "◆". These graphs were created using the program GraPhlAn (Asnicar *et al.*, 2015).

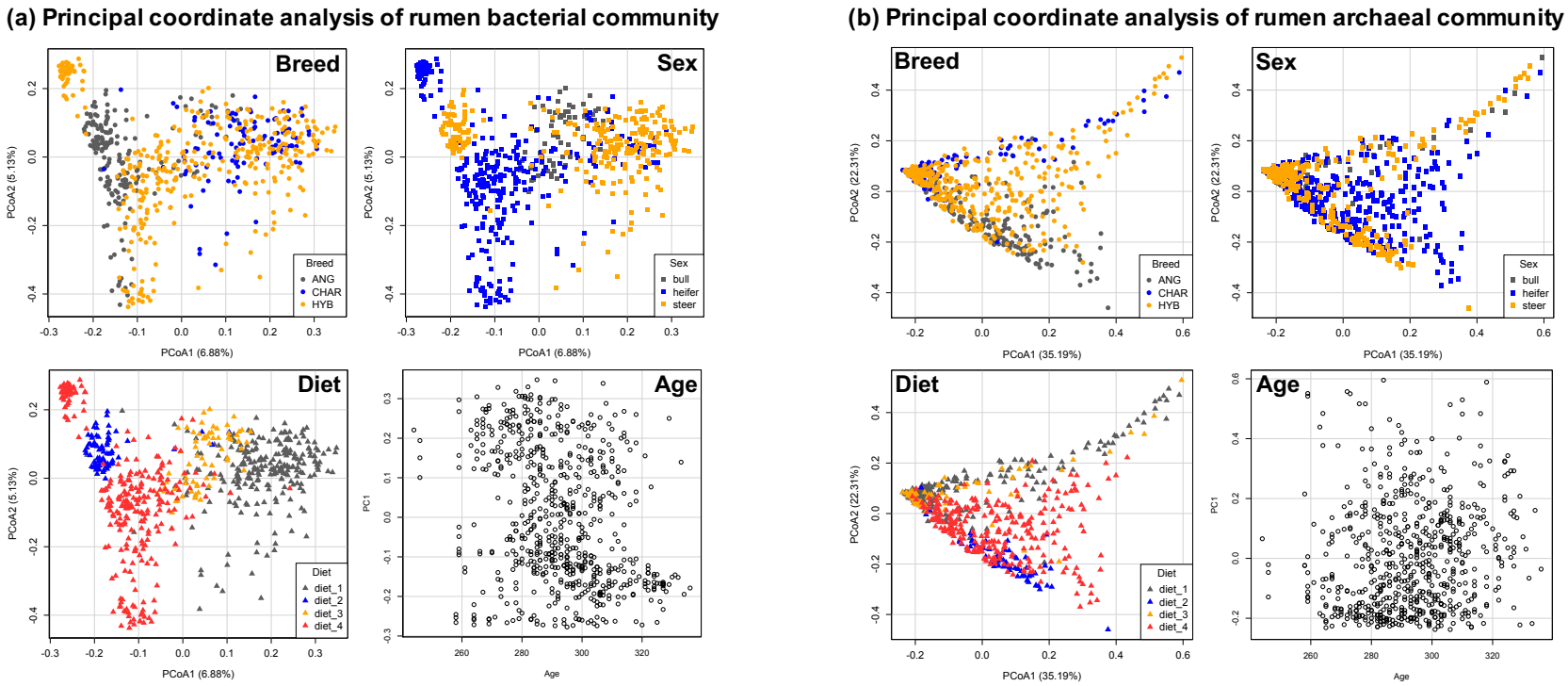


Figure 5.2 Factors (breed, sex, diet, and age) drive segregation of rumen bacterial communities (**a**) and archaeal communities (**b**), as visualized using principal coordinate analysis (PCoA). To performed PCoA, the number of bacterial and archaeal sequences per sample were normalized to 2,000 and 500, respectively, and the PCoA was conducted using Bray-Curtis dissimilarity matrices.

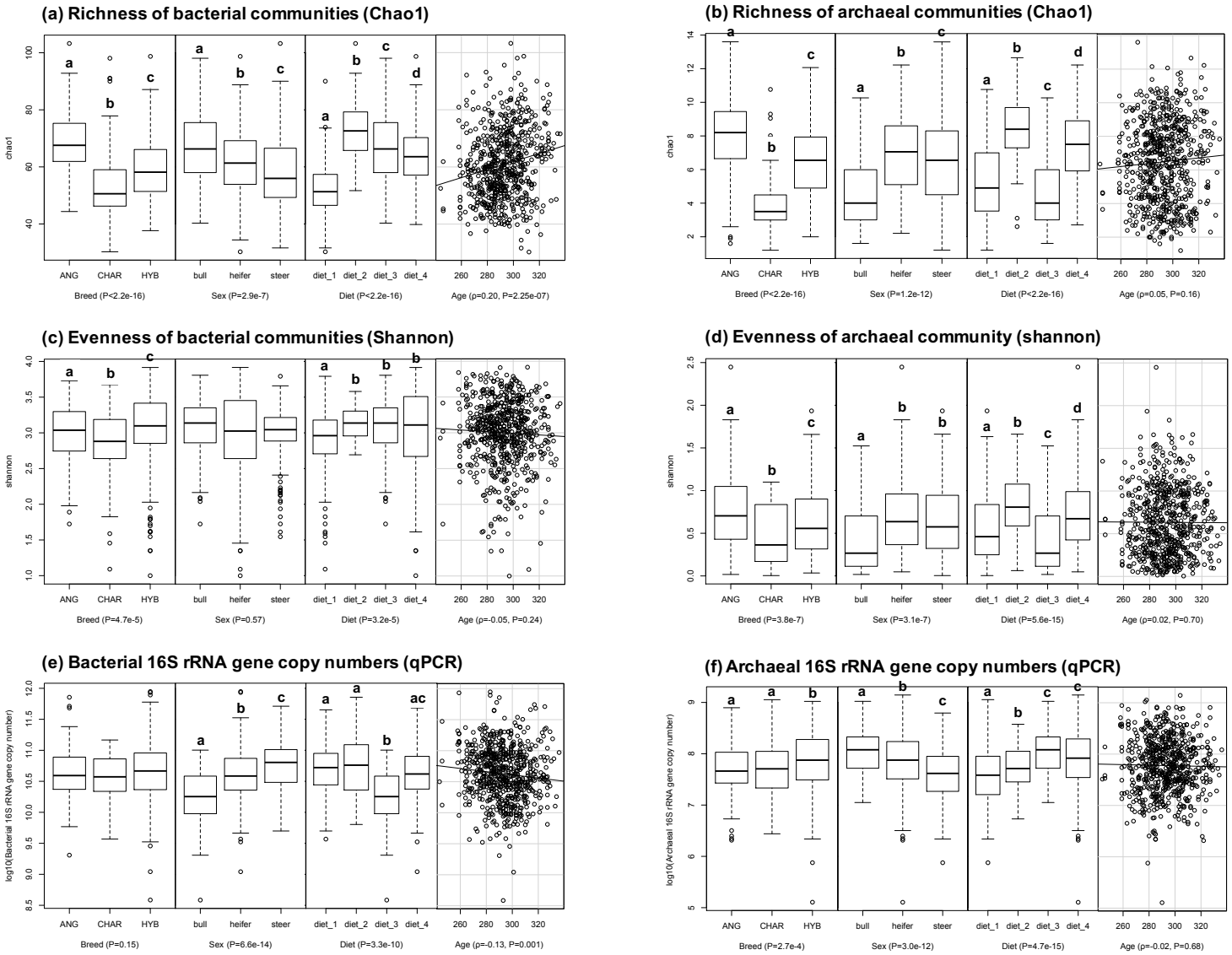


Figure 5.3 Effects of breed, sex, diet, and age on the alpha-diversities and abundance of rumen bacteria and archaea. The 16S rRNA gene copy numbers per ml of rumen sample were log10-transformed before statistical analysis. Values within each factor that do not have a common superscript are significantly different ($P < 0.05$) according to the Kruskal-Wallis rank sum test. The correlations between age and other indices were calculated using the Spearman's rank correlation (ρ = correlation coefficient).

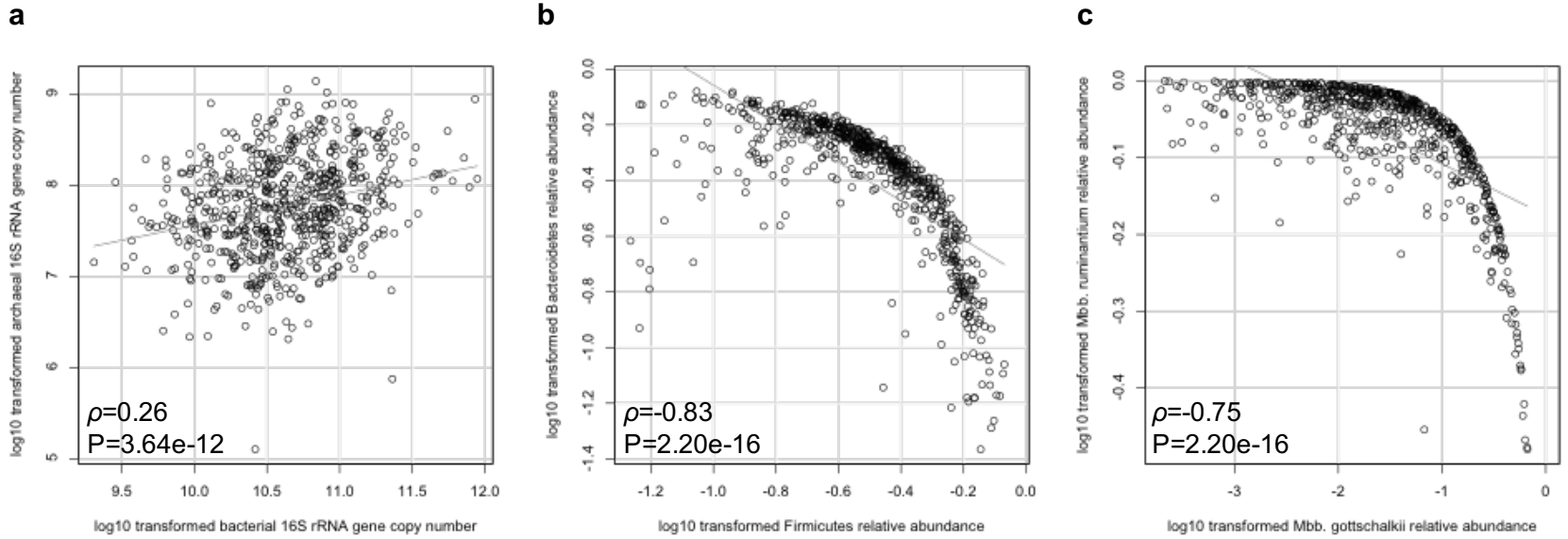


Figure 5.4 Relationships between predominant rumen microbial groups. (**a**) Ratio of bacterial abundance to archaeal abundance, represented by 16S rRNA gene copy number obtained using qPCR. (**b**) Ratio of *Firmicutes* to *Bacteroidetes*. (**c**) Ratio of *Methanobrevibacter gottschalkii* to *Methanobrevibacter ruminantium*. The 16S rRNA gene copy number and relative abundance were log10-transformed, and the correlation analysis was performed using the Spearman's rank correlation (ρ = correlation coefficient).

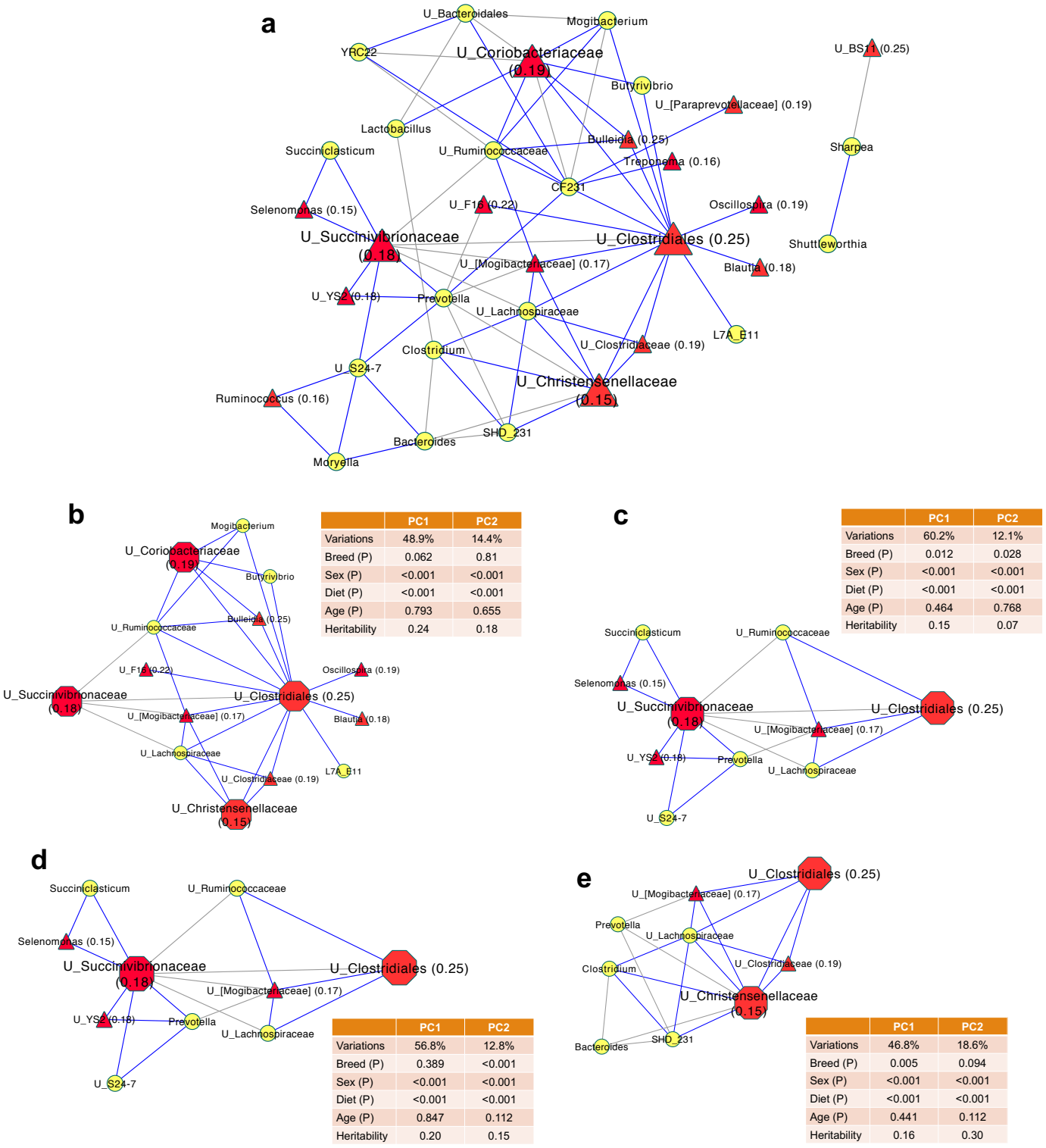
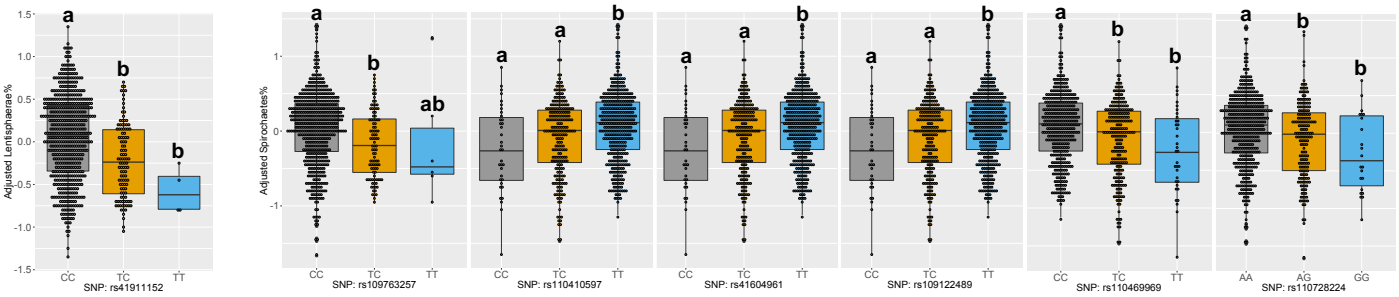
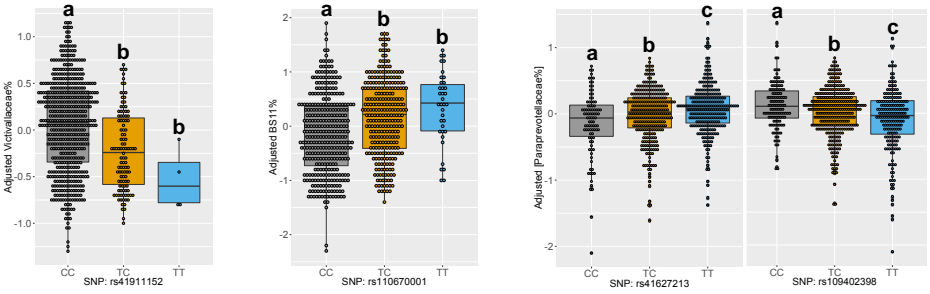


Figure 5.5 Co-occurrence network of rumen microbial taxa (**a**). Four major co-occurrence network modules were centered by unclassified *Clostridiales* (**b**), unclassified *Succinivibrionaceae* (**c**), unclassified *Coriobacteriaceae* (**d**), and unclassified *Christensenellaceae* (**e**). Only correlations with coefficient > 0.3 or < -0.3 and with *P* value < 0.001 were displayed. Heritable taxa were represented by red triangle, while inheritable taxa were represented by yellow circle. Values in the parentheses are heritability estimates of heritable taxa. A connection with a blue/grey line means a positive/negative correlation. '*U_*' before the taxonomic name represents unclassified. The first two PCs were calculated using PCA for each module.

(a) SNPs associated with phylum-level taxa



(b) SNPs associated with family-level taxa



(c) SNPs associated with genus-level taxa

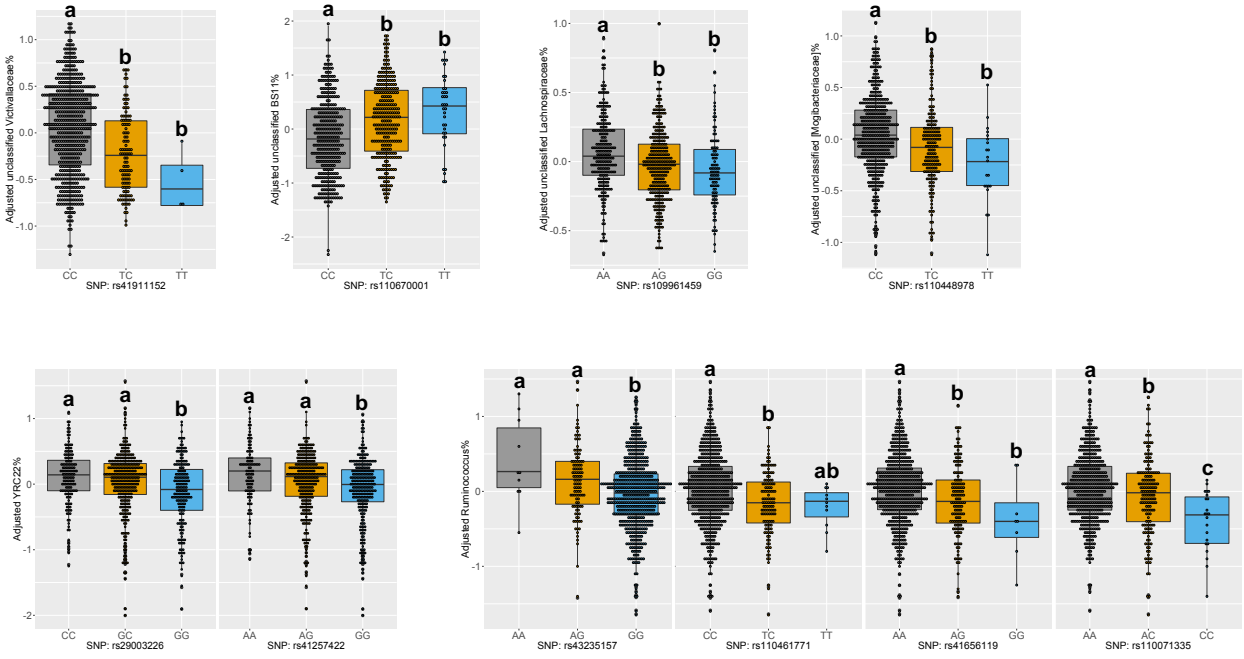


Figure 5.6 SNPs associated with rumen microbial taxa (at phylum, family, and genus levels). Only associations with false discovery rates (FDR) < 0.1 (significant) and 0.1 < FDR < 0.2 (suggestively significant) are displayed. In each plot, values that do not have a common superscript are significantly different ($P < 0.05$) based on ANOVA. The x axis represents genotype of a SNP, and the y axis indicated the log10-transformed relative abundance after adjusting breed, sex, diet, and age factors.

Chapter 6. General discussion

Feed efficiency is one of determinants of the sustainability of the beef cattle production. In the past, many effects have been made to improve feed efficiency through breeding and genetics, optimal feeding, and improved management. To date, evidence is accumulating regarding the associations between feed efficiency and rumen microorganisms in ruminants. However, our understanding of the rumen microbiome and its linkages to host feed efficiency is still limited, due to 1) the lack of reliable methods to characterize rumen microbial activities and functions, 2) the lack of capacity to collect high amount of rumen samples from the real and large beef population, and 3) without the knowledge of the host genetic effect on the rumen microbiome.

Findings in this thesis have broadened our knowledge on the rumen microbiome and its functions in beef cattle. Firstly, an experimental and bioinformatic pipeline was developed to identify and quantify the active rumen microbiota using total-RNA-based metatranscriptomic analysis (Chapter 2). Secondly, rumen microbial taxonomic and functional features were characterized using metatranscriptomics, and then linked to feed efficiency in beef cattle from a single breed (Kinsella composite hybrid; Chapter 3). Thirdly, associations between the rumen microbiome and feed efficiency were identified for beef cattle from three different breeds (Angus, Charolais, and Kinsella composite hybrid), at both DNA and RNA levels using metagenomics and metatranscriptomics (Chapter 4). Furthermore, through a genome-wide association analysis (GWAS) for a large cohort of beef population in Chapter 5, moderate heritabilities were detected for the diversity of rumen microbiota as well as ~30% of detected microbial taxa, and a total of 19 SNPs across 12 *Bos taurus* autosomes were identified to be

associated with relative abundance of 14 microbial taxa. Therefore, this thesis filled several knowledge gaps on the rumen microbiome, and provided a basis for the further improvement of feed efficiency through manipulating the rumen microbiome in beef cattle. Main findings, limitations, and future directions are discussed in following sections.

6.1 Active rumen microbiota estimated using total-RNA-based metatranscriptomics

Compared with previous studies, the developed pipeline to estimate the active rumen microbiota (Chapter 2) was based on two custom-built 16S rRNA gene regionally enriched references: bacterial V1-V3 regional database (based on Greengenes 16S rRNA gene database) (DeSantis *et al.*, 2006) and archaeal V6-V8 regional databases (based on rumen-specific archaeal 16S rDNA database, or Rumen and Intestinal Methanogen-DB [RIM-DB]) (Janssen and Kirs, 2008; Seedorf *et al.*, 2014). Considering the short length of Illumina sequences and the existence of conserved regions in 16S rRNA genes, if short rRNA reads are directly aligned to full-length 16S rRNA gene databases as previous studies did (Tveit *et al.*, 2014; Urich *et al.*, 2008), there would be multiple hits, and thus mixed taxonomic profiles and biased estimation of diversity. These two regional enriched references were generated to overcome this shortcoming. Future studies are needed to further validate the accuracy of our developed pipeline, using a mock microbial community, consisting of mixed phylotypes with known composition and genomes.

From total-RNA-based metatranscriptomes, it is notable that a large fraction of rRNA was classified as eukaryotic 18S and 28S rRNA. Although these reads were not analyzed in this thesis, the high number of these sequences indicates the possibility of assessing rumen eukaryotes using RNA-seq in future studies. We have made some efforts to generate the taxonomic profiles for protozoa and fungi through the enrichment of 18S rRNA, and/or ITS

reads from total RNA. However, due to limited databases for rumen eukaryotes, the assessment of ciliates and fungi are not convincing.

In contrast to the predominance of *Bacteroidetes* and *Firmicutes* reported in previous rumen studies (Henderson *et al.*, 2015; Kim *et al.*, 2011), this thesis revealed *Proteobacteria* was the most abundant bacterial phylum in the rumen of steers at the RNA level (Chapters 2 and 3). However, when more metatranscirptomes (n = 48) were analyzed using the same pipeline for taxonomic assessment (Chapter 4), *Bacteroidetes* and *Firmicutes* were more abundant than *Proteobacteria* in those cattle, which is in line with our common knowledge on the rumen microbial composition. These differences may be caused by the different storage conditions between those rumen samples. Samples for Chapters 2 and 3 were stored in RNA*later* at -20 °C, while samples for Chapter 4 were snap-frozen using liquid nitrogen and stored under -80°C. Although it has been suggested that there is no significant difference between RNA*later* and liquid nitrogen freezing methods for human gut metatranscriptomes (Franzosa *et al.*, 2014), no study has been done to compare the effects of these two storage methods on rumen metatranscriptomic profiles. Therefore, it is necessary to perform further comparison studies to evaluate the storage effects on rumen metatranscriptomic profiling in the future. In addition, rumen samples in Chapters 2 and 3 were collected in 2007 and were sequenced in 2013, whiles samples in Chapters 4 and 5 were collected in 2014 and were sequenced in 2015. These different storage times may also contribute to these observed differences as a recent study indicated (Granja-Salcedo *et al.*, 2017).

According to results of Chapters 2-4, *Methanomassiliicoccales* (MCC) is the most active archaeal group in rumen (35 - 42%), which was previously called Rumen Cluster C (RCC) or *Thermoplasmatales* (Gaci *et al.*, 2014; Janssen and Kirs, 2008; Poulsen *et al.*, 2013). Early

studies at the DNA level reported that the abundance of MCC was around 16% and usually less than 20% (Henderson *et al.*, 2015; Janssen and Kirs, 2008; Kittelmann *et al.*, 2013), lower than *Methanobrevibacter gottschalkii* and *Methanobrevibacter ruminantium*. However, high abundance of MCC at the RNA level indicates that members of this group may play more important and essential roles in rumen methanogenesis in beef cattle than we thought. It is known that methanol and methylamines are major energy and carbon sources for MCC (Poulsen *et al.*, 2013), and thus the methyl compound conversion pathway is probably the primary methanogenesis route in rumen of beef cattle. Transcripts of MCC were directly associated with CH_4 emissions (Poulsen *et al.*, 2013) and feed efficiency in cattle (Chapters 3 and 4). Till now, only a few members of MCC have been isolated and have available genomic information: most of them are human isolates (Noel *et al.*, 2016) and only two are from rumen isolates (Li *et al.*, 2016; Noel *et al.*, 2016). Therefore, to better understand their ecological niches and functions in rumen, more pure-culture based effects should be made to study members of MCC. Furthermore, due to its associations with CH_4 emissions and feed efficiency, MCC could be considered as one of targeted taxonomic groups to manipulate for improved RFI and decreased CH_4 emissions.

6.2 Understanding of rumen microbial functional activities in beef cattle

In this thesis, it was observed that rumen microbiomes had higher functional variations at the RNA level than at the DNA level (Chapter 4), and this pattern was also reported in microbiomes of human gut (Franzosa *et al.*, 2014; Nayfach *et al.*, 2015) and soil (Urich *et al.*, 2008). This indicates there may be a similar functional potential pool existing in rumen microbiomes across all beef cattle under the same dietary and environmental conditions. Indeed, all 48 steers had the same diet and were raised under the same condition. Due to this similarity,

the real linkages between the rumen microbiome and host phenotypes may not be detected using metagenomics. For example, Shi *et al.* (2014) identified rumen microbial gene expression levels (transcripts abundance) were more related to methane emissions rather than gene abundance. Furthermore, several studies have suggested that gene expression profiles of gut microbiomes varied significantly under dietary changes but community structures did not change significantly (Maurice *et al.*, 2013; McNulty *et al.*, 2011). Collectively, it can be speculated that rumen microbial functional activities, rather than microbial community structures and functional gene abundance, contribute to host phenotypic variations.

Therefore, to better understand functional activities of rumen microbiomes, this thesis has focused on microbial functional profiles in the rumen of beef cattle through the metatranscriptomic analysis (Chapters 3 and 4). In addition to metatranscriptomics, metaproteomics is an alternative approach to generate active microbial functional profiles. Compared to RNA, proteins have longer half-life (Hargrove and Schmidt, 1989) and higher abundance in cells (Schwanhausser *et al.*, 2011). What's more, metatranscriptomics does not account for the dynamics between protein synthesis and degradation as well as post translational modification, and not all transcripts identified from metatranscriptomes are translated to proteins that directly involve in various biological processes. Accordingly, metaproteomics has been suggested as a more accurate method to reveal true functions than metatranscriptomics (Haange and Jehmlich, 2016). Therefore, future studies to involve metaproteomic analysis may improve our understanding of rumen microbial functions. However, due to the lack of characterized rumen microbial proteins, metaproteomics is still challenging. For example, in a recent study aiming to profile rumen metaproteomes, only a few hundred proteins were identified (Snelling and Wallace, 2017). Moreover, as an extension study of this thesis, metabolomic analysis is

under way for rumen digesta samples that have been studied using metagenomics and metatranscriptomics in Chapter 4, and these global products/metabolites could be used as direct evidence of rumen microbial functional activities.

6.3 Associations between the rumen microbiome and feed efficiency in cattle

Early studies have suggested that there are associations between taxonomic characteristics (e.g., abundance and/or prevalence of microbial taxa, diversity indices, etc.) and cattle feed efficiency (summarized in **section 1.6.1**), but no obvious taxonomic difference was detected between high (L-RFI) and low (H-RFI) feed efficiency steers in Chapter 3: with only four microbial taxa tending to be different between H- and L-RFI animals ($0.05 < P < 0.1$). The lack of difference is because the analysis in Chapter 3 was conducted at the family level for bacteria. It is possible that different bacterial genera within the same family have different ecologic niches, while bacteria taxa belonging to different families share similar ecologic niches (Jami and Mizrahi, 2012; Klieve *et al.*, 2007). This probably has masked the actual linkages between microbial taxa and feed efficiency. Therefore, the analysis in Chapter 4 was moved to deeper taxonomic levels (to the genus level for bacteria and to the species level for archaea), to make sure each microbial taxonomic group has more similar niches. As expected, there are more bacterial taxa at the genus level showed different abundance between feed efficiency groups. Accordingly, to better identify the associations between rumen microbial taxa and feed efficiency, it is better to conduct the analysis at deeper taxonomic levels (e.g., genus, species, strains, etc.) rather than at higher taxonomic levels (e.g. phylum, family, and so on).

In Chapters 3 and 4, t-test was used to detect differential microbial taxa between high and low RFI groups based on the arcsine-square-root-transformed relative abundance (proportion) as

previously suggested (Franzosa *et al.*, 2014). This data transformation was conducted to stabilize the variance of the proportion data, making them more appropriate for frequently-used normal-distribution-based statistical methods. However, it is worth mentioning that the transformed relative abundance may still not follow the normal distribution, and thus does not fully satisfy the assumption of t-test, especially for those taxa with low abundance and low prevalence. Therefore, in order to more accurately identify differential microbial taxa between RFI groups, several programs and/or methods should be considered in the future, such as LEfSe (Segata *et al.*, 2011), DESeq2 (Love *et al.*, 2014), edgeR (Robinson and Smyth, 2007), and ANCOM (Mandal *et al.*, 2015). A recent study compared the accuracy of several widely-used programs/methods (including DESeq2, edgeR, ANCOM, and so on) for the comparison analysis (Weiss *et al.*, 2017). It reported that DESeq2 was sensitive for small datasets (< 20 samples per group), while ANCOM could maintain a low false discovery rate but it was not sensitive for small datasets (< 20 samples per group). According to the sample size in Chapters 3 and 4, DESeq2 could be a more accurate option to detect differential microbial phylotypes between RFI groups.

Furthermore, recent studies indicate that feed efficiency is actually more related to rumen microbial functional features rather than taxonomic features (Roehe *et al.*, 2016; Shabat *et al.*, 2016). In Chapters 3 and 4, most differential microbial genes/transcripts/pathways between H- and L-RFI groups were enriched in inefficient individuals than in efficient ones. This supports recent findings using metagenomics (Shabat *et al.*, 2016), where rumen microbiomes of inefficient cattle had more diverse functional potentials than those of feed efficient cattle. All these suggest that rumen microbiomes of inefficient animals may possess larger capacities and higher activities to utilize a wider range of substrates, and thus to generate higher amount and more diverse products. On the one hand, the amount of these products may exceed the maximum

absorptive capacity of hosts, leading to either harmful accumulation or energy waste, and thus lower feed efficiency (Shabat *et al.*, 2016). On the other hand, higher activities of rumen microbiomes in inefficient cattle imply higher ratio of microbial proliferation, and thus microorganisms themselves also consumed more nutrients/substrates. In contrast, rumen microbiomes in efficient cattle may have relatively conserved functional potentials and more appropriate activities, which could generate more relevant products and effectively balance the rumen microbial fermentation and the host absorption. Currently metabolomic analysis of these rumen digesta samples and qPCR of rumen bacteria and archaeal 16S rRNA genes are under way, and outcomes of products/metabolites plus abundance estimation for rumen microbiomes could tell us whether these speculations are true.

More differential abundant functional categories between feed efficiency groups were detected at the RNA level than at the DNA level (11 versus 3; Chapter 4), suggesting that feed efficiency is more affected by activities of rumen microbiomes and less affected by their functional genetic potentials. Moreover, many differential abundant genes between feed efficiency groups were found from metagenomes, but most of them did not show different expression levels from metatranscriptomes (Chapter 4), which represent potential targets for further exploration. Working on these differential abundant genes to alter their expression levels could be a possible strategy to manipulate the rumen microbiome for improved feed efficiency in beef cattle.

In addition, most of differential microbial taxa and functional features between feed efficiency groups only show differences within one breed, rather than shared among all three breeds (Chapter 4). This suggests that feed efficiency of beef cattle should be partially affected by the host and microbiome interactions, rather than the microbiome itself. This could explain

the generally low consistence among different studies when rumen microbiomes were linked to feed efficiency.

6.4 Rumen microbiome is affected by host genetics

The observed breed-associated variations of the rumen microbiome (Chapter 4) represent potential genetic superiorities of each breed. For example, among three breeds (Angus [ANG], Charolais [CHAR], and Kinsella composite hybrid [HYB]), CHAR had the simplest rumen microbial communities. It was suggested that a simpler rumen microbial community was associated with higher feed efficiency and lower CH_4 emissions (Shabat *et al.*, 2016; Zhou *et al.*, 2009). Indeed, a previous study found that CHAR steers consumed ~5% less daily dry matter intake (DMI) and had 2.7% more average dairy gain (ADG) than ANG steers (Mao *et al.*, 2013). Therefore, crossing CHAR with other cattle breeds is a feasible strategy to reduce the complexity of rumen microbial communities in other breeds, and thus to enhance feed efficiency and reduce CH_4 yields in the future. Meanwhile, around half of detected rumen bacterial and archaeal taxa showed differential abundance among three breeds, indicating the possibility to manipulate particular rumen microbial taxa using crossbreeding. This possibility is becoming more and more meaningful, because currently an increased number of host-phenotype-associated rumen microbial taxa are being identified, and these host-phenotype-associated microbial taxa are potential targets for further improving host phenotypes.

Furthermore, the obtained moderate heritabilities of rumen microbiota provide the first evidence that rumen microbial communities are partially shaped by host additive genetic effects, and it is the first time to link rumen microbial profiles and host genotypes in beef cattle (Chapter 5). Therefore, manipulating heritable rumen microorganisms using quantitative genetics becomes

more practicable. These heritable microbial features could be regarded as breeding objectives and be involved to the breeding program, which could be ultimately used to improve rumen microbiota-associated traits, such as feed efficiency, CH_4 emissions, and so on. Through the GWAS, several rumen microbiota-associated SNPs were found to overlap with known QTLs for feed efficiency in cattle, such as average daily gain (ADG) (de Oliveira *et al.*, 2014; Li *et al.*, 2002; Rolf *et al.*, 2012) and residual feed intake (RFI) (Sherman *et al.*, 2009). These QTLs may have pleiotropic effects on both the rumen microbiota and feed efficiency, which may explain the associations between rumen microorganisms and feed efficiency. Meanwhile, it is also possible that host QTLs control feed efficiency through their effects on altering the rumen microbial composition. Because VFAs produced by the rumen microbiota could meet 70% of the daily energy requirement for the host (Bergman, 1990), they are considered as important intermedia between the host and the rumen microbiota that affect feed efficiency and CH_4 emissions (Guan *et al.*, 2008; Shabat *et al.*, 2016). Therefore, to further understand how the rumen microbiota influences feed efficiency and CH_4 emissions, rumen VFA profiles of the same animals have been generated and will be used to predict CH_4 emissions (Angela *et al.*, 2000). These data are not included in this thesis because their analysis is part of another PhD student's thesis work. In the future, the linkages among VFAs, microbiota, and host genetics can provide more conclusive understanding on the role of host-microbial interactions in cattle feed efficiency. Further analyses to explore phenotypic and genetic correlations among the rumen microbiota, VFA profiles, feed efficiency, and CH_4 emissions may reveal the mechanisms behind their relationships, which could help us better define strategies to improve feed efficiency and decrease CH_4 emissions through working on the rumen microbiota.

6.5 Limitations

6.5.1 Rumen sampling could impact the outcomes of microbial profiling

Firstly, rumen digesta samples used in Chapters 2-4 were collected from steers at slaughter, and it has been suggested that rumen microbial profiles were different between samples taken from live animals and at slaughter (Wallace *et al.*, 2014). Although Wallace *et al.* (2014) considered the archaeal abundance in post-mortem rumen digesta as a proxy measurement for methane emissions for steers, rumen samples collected from live animals may still be more appropriate for rumen microbial profiling and for linking them with host phenotypes. Secondly, all rumen samples used for this thesis (Chapters 2-5) were collected in the morning before feeding. Li *et al.* (2009b) did not find obvious effects of sampling time on the rumen bacteria using PCR-DGGE and qPCR, but Golder *et al.* (2014) reported that the rumen bacterial community shifted at different sampling times (5, 115, 215 mins after feeding) using 16S rRNA gene amplicon sequencing. Sampling time may have more significant effects on metatranscriptomes than DNA-based outcomes, because gene expressions of microbial genes may shift over time due to the amount of available substrates without necessarily changing the microbial composition and functional potentials (Firkins and Yu, 2015). Thirdly, samples for Chapter 5 were rumen fluid collected through oro-gastric tubing, which is a practical sampling method for large number of animals. However, generating microbial profiles using rumen fluid samples ignored the bacterial population that attached to feed particles (70-80% of total rumen bacteria) (Minato *et al.*, 1993). Additionally, because sampling was conducted for a real and large beef population, it is impossible to collect rumen samples multiple times within one day or within a short period. In that case, this one-time sampling may bring biases to the estimation of microbial profiles. Fourthly, for the oro-gastric tubing, it has been reported that the insertion

depth of tubes affected the fermentation parameters of rumen fluid (Shen *et al.*, 2012), and thus it may also influence the rumen microbial composition. Therefore, to more accurately generate rumen microbial profiles and link them with host performances, all these variables related to rumen sampling should be considered and controlled in future studies.

6.5.2 Rumen sampling under different diets

It has been reported that beef steers changed their RFI classifications when their diets were switched from the low energy (growing) diet to the high energy (finishing) diet (Durunna *et al.*, 2011), and there were different rumen microbial communities under these two types of diets (Ellison *et al.*, 2014; Petri *et al.*, 2013). Meanwhile, our previous study already revealed that there were different rumen bacteria phylotypes contributing to the RFI variations under two different diets (Hernandez-Sanabria *et al.*, 2012). In the beef cattle operations, most of feed costs and grain consumption were attributed to the finishing diet. Therefore, to improve the feed efficiency and reduce the costs of feed, identifying microbial features associated with RFI under the finishing diet is more economically meaningful than identify them under the growing diet. That is the reason that we used rumen samples collected at the finishing stage of beef steers in Chapters 2-4. However, heritabilities of rumen microbial features were estimated based on rumen samples collected under the low-energy diet rather than the high-energy diet in Chapter 5. Because in Chapter 5, we included a large cohort of beef population consisting of three genders of beef cattle (bull [n = 71], heifer [n = 350], and steer [n = 291]) from the same generation to increase the accuracy of heritability estimation, and it is not realistic to feed bulls and heifers with high-energy diet in the real beef cattle operations. Therefore, although rumen samples were also collected from steers under the finishing diet (n = 261), the estimation of heritability may be

less accurate due to the small population size. To better detect the heritable rumen microbial taxa under the finishing diet, rumen samples of steers under the finishing diet were already collected from another two generations, and further analysis will be conducted to evaluate the host genetic effect on the rumen microbiota under the high energy diet.

6.5.3 Comparison between total-RNA-based and mRNA-enriched metatranscriptomes

According to the current comparison analysis between total-RNA-based and mRNA-enriched metatranscirptomes (Chapter 4), we suggested that the mRNA enrichment led to the underestimation of many expressed genes and biased the abundance estimation of certain functional categories. It may be true, because the mRNA enrichment may bring about biases caused by the mRNA degradation during the extended sample processing time. But it is also possible that these results and conclusions are inappropriate due to the analysis strategy that we applied: in the current analysis, relative abundance (proportion) of transcripts was compared and only transcripts with a minimum relative abundance of 0.01% in at least three samples within a dataset remained for the comparison analysis as suggested by Franzosa *et al.* (2014). This comparison analysis based on relative abundance is actually not fair for mRNA-enriched metatranscirptomes, because they had almost 20 times more annotated mRNA reads than total-RNA-based metatranscirptomes. In that case, differences caused by different sequencing depths were overlooked. Therefore, to further check the effect of the mRNA enrichment and conduct more comprehensive comparison analysis, it may be better to compare the absolute read counts directly in future.

Our results have showed that the mRNA enrichment significantly increased the number of mRNA reads (~9 folds) and the number of annotated mRNA reads (~20 folds), compared to

total-RNA-based metatranscriptomes. It suggests that the enrichment of mRNA successfully increased the sequencing depth of mRNA, which is the advantage of mRNA-enriched metatranscriptomics. However, whether the mRNA enrichment is necessary depends on the research questions: if the main objective is to explore specific transcripts/pathways, to capture more transcripts, or to generate more comprehensive gene expression profiles, the mRNA enrichment is suggested; if researchers aim to generate both taxonomic and functional profiles simultaneously, total-RNA-based metatranscirptomics should be applied, because rRNA and mRNA could be used to estimate the composition and functions of the rumen microbiome, respectively.

6.7 Implications

In Chapters 2 and 3, a pipeline was developed to explore taxonomic and functional characteristics of the rumen microbiome using metatranscriptomics. This adds a new alternative approach to existing microbiological methods, which allows us to obtain more real-time estimations for the alterations of microbial population and functional activities in rumen. An accurate evaluation of microbial profiles is crucial for further associating the rumen microbiome to host phenotypes. In Chapter 4, rumen metagenomes and metatranscriptomes were compared, and separated functional profiles were identified between DNA and RNA levels. This observation implied that previously reported associations between the rumen microbiome at the DNA level and host phenotypes may not represent their actual linkages. To build more solid and accurate associations between the rumen microbiome and feed efficiency or other host phenotypes, metatranscriptomics is recommended to profile the rumen microbiome at both taxonomic and functional levels. At the same time, the comparison between total-RNA-based

and mRNA-enriched metatranscriptomes suggested that whether including the mRNA enrichment should depend on research questions. In Chapter 5, heritable rumen microbial taxa in rumen were detected, and bovine genetic components associated with rumen microbial taxa were identified. These findings provide the evidence that the rumen microbiota is partially shaped by host genetics.

In addition to novel methods and fundamental knowledge on the rumen microbiome, outcomes of this thesis can also be applied to the industry as well. Identified feed-efficiency-associated rumen microbial taxa and functional features could be useful markers to predict the feed efficiency performance in beef cattle (Chapters 3 and 4). Nowadays, to calculate RFI, individual feed intake must be recorded for 9-12 weeks using the automatic feeding measurement system, which is time-consuming and expensive. Our results suggest the possibility to design the microbial array that contains probes for these predictive microbial markers of RFI (such as marker genes for microbial abundance/activities and functional genes for gene abundance/expressions). It will increase the feasibility to quickly generate microbial profiles with low costs using rumen samples and thus to predict the RFI performance using microbial markers. Because rumen samples could be collected at early time point, this strategy could significantly reduce the time of RFI measurement, as well as the costs associated with the measurement. Furthermore, moderate heritability estimates were detected for several feed efficiency and/or CH_4 emissions related rumen microbial features, and several candidate genes including *RAPH*1, *CDC*7, and *MYH*3 were suggested to be associated with the rumen microbiota. Several rumen microbiota-associated SNPs were found to overlap with known QTLs for feed efficiency in cattle. These findings provide an initial but prospective foundation for the further genomic selection of a desirable rumen microbiota corresponding to high productivity

and low environmental impact. For example, *Blautia* showed higher abundance in L-RFI individuals than in H-RFI ones (Chapter 4). Considering this group had a heritability of 0.18 (Chapter 5), its abundance could be elevated through selecting and breeding individuals with high abundance of *Blautia*, which could be served as a potential way to improve feed efficiency.

6.8 References

Angela RM, Jean-Pierre J, John N (2000). Methane production by ruminants: its contribution to global warming. *Ann Zootech* **49:** 231-253.

Bergman EN (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev* **70:** 567-590.

de Oliveira PS, Cesar AS, do Nascimento ML, Chaves AS, Tizioto PC, Tullio RR *et al* (2014). Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genet* **15:** 100.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069-5072.

Durunna ON, Mujibi FD, Goonewardene L, Okine EK, Basarab JA, Wang Z *et al* (2011). Feed efficiency differences and reranking in beef steers fed grower and finisher diets. *J Anim Sci* **89:** 158-167.

Ellison MJ, Conant GC, Cockrum RR, Austin KJ, Truong H, Becchi M *et al* (2014). Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. *DNA Res* **21:** 115-125.

Firkins JL, 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:** 1450-1470.

Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM *et al* (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* **111:** E2329-2338.

Gaci N, Borrel G, Tottey W, O'Toole PW, Brugere JF (2014). Archaea and the human gut: new beginning of an old story. *World J Gastroenterol* **20:** 16062-16078.

Golder HM, Denman SE, McSweeney C, Celi P, Lean IJ (2014). Ruminal bacterial community shifts in grain-, sugar-, and histidine-challenged dairy heifers. *J Dairy Sci* **97:** 5131-5150.

Granja-Salcedo YT, Ramirez-Uscategui RA, Machado EG, Duarte Messana J, Takeshi Kishi L, Lino Dias AV *et al* (2017). Studies on bacterial community composition are affected by the time and storage method of the rumen content. *PLoS One* **12:** e0176701.

Guan LL, Nkrumah JD, Basarab JA, Moore SS (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol Lett* **288:** 85-91.

Haange SB, Jehmlich N (2016). Proteomic interrogation of the gut microbiota: potential clinical impact. *Expert Rev Proteomics* **13:** 535-537.

Hargrove JL, Schmidt FH (1989). The role of mRNA and protein stability in gene expression. *Faseb j* **3:** 2360-2370.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH (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.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Durunna ON, Moore SS, Guan LL (2012). Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. *Appl Environ Microbiol* **78:** 1203-1214.

Jami E, Mizrahi I (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS One* **7:** e33306.

Janssen PH, Kirs M (2008). Structure of the archaeal community of the rumen. *Appl Environ Microbiol* **74:** 3619-3625.

Kim M, Morrison M, Yu Z (2011). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol Ecol* **76:** 49-63.

Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI *et al* (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS One* **8:** e47879.

Klieve AV, O'Leary MN, McMillen L, Ouwerkerk D (2007). Ruminococcus bromii, identification and isolation as a dominant community member in the rumen of cattle fed a barley diet. *J Appl Microbiol* **103:** 2065-2073.

Li C, Basarab J, Snelling WM, Benkel B, Murdoch B, Moore SS (2002). The identification of common haplotypes on bovine chromosome 5 within commercial lines of Bos taurus and their associations with growth traits. *J Anim Sci* **80:** 1187-1194.

Li M, Penner GB, Hernandez-Sanabria E, Oba M, Guan LL (2009). Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen. *J Appl Microbiol* **107:** 1924-1934.

Li Y, Leahy SC, Jeyanathan J, Henderson G, Cox F, Altermann E *et al* (2016). The complete genome sequence of the methanogenic archaeon ISO4-H5 provides insights into the methylotrophic lifestyle of a ruminal representative of the Methanomassiliicoccales. *Stand Genomic Sci* **11:** 59.

Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.

Mandal S, Van Treuren W, White RA, Eggesbo M, Knight R, Peddada SD (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis* **26:** 27663.

Mao F, Chen L, Vinsky M, Okine E, Wang Z, Basarab J *et al* (2013). Phenotypic and genetic relationships of feed efficiency with growth performance, ultrasound, and carcass merit traits in Angus and Charolais steers. *J Anim Sci* **91:** 2067-2076.

Maurice CF, Haiser HJ, Turnbaugh PJ (2013). Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* **152:** 39-50.

McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL *et al* (2011). The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci Transl Med* **3:** 106ra106.

Minato H, Mitsumori M, Cheng KJ (1993). Attachment of microorganisms to solid substrates in the rumen. *Genetics, Biochemistry and Ecology of Lignocellulose Degradation (Shimada, K, Ohmiya, K, Kobayashi, Y, Hoshino, S, Sakka, K and Karita, S, Eds)***:** 139-145.

Nayfach S, Bradley PH, Wyman SK, Laurent TJ, Williams A, Eisen JA *et al* (2015). Automated and Accurate Estimation of Gene Family Abundance from Shotgun Metagenomes. *PLoS Comput Biol* **11:** e1004573.

Noel SJ, Hojberg O, Urich T, Poulsen M (2016). Draft Genome Sequence of "Candidatus Methanomethylophilus" sp. 1R26, Enriched from Bovine Rumen, a Methanogenic Archaeon Belonging to the Methanomassiliicoccales Order. *Genome Announc* **4**.

Petri RM, Schwaiger T, Penner GB, Beauchemin KA, Forster RJ, McKinnon JJ *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:** e83424.

Poulsen M, Schwab C, Jensen BB, Engberg RM, Spang A, Canibe N *et al* (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat Commun* **4:** 1428.

Robinson MD, Smyth GK (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23:** 2881-2887.

Roehe R, Dewhurst RJ, Duthie CA, Rooke JA, McKain N, Ross DW *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 Genet* **12:** e1005846.

Rolf MM, Taylor JF, Schnabel RD, McKay SD, McClure MC, Northcutt SL *et al* (2012). Genome-wide association analysis for feed efficiency in Angus cattle. *Anim Genet* **43:** 367-374.

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J *et al* (2011). Global quantification of mammalian gene expression control. *Nature* **473:** 337-342.

Seedorf H, Kittelmann S, Henderson G, Janssen PH (2014). RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. *PeerJ* **2:** e494.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS *et al* (2011). Metagenomic biomarker discovery and explanation. *Genome Biol* **12:** R60.

Shabat SK, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Berg Miller ME *et al* (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *Isme j*.

Shen JS, Chai Z, Song LJ, Liu JX, Wu YM (2012). Insertion depth of oral stomach tubes may affect the fermentation parameters of ruminal fluid collected in dairy cows. *J Dairy Sci* **95:** 5978-5984.

Sherman EL, Nkrumah JD, Li C, Bartusiak R, Murdoch B, Moore SS (2009). Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. *J Anim Sci* **87:** 37-45.

Shi W, Moon CD, Leahy SC, Kang D, Froula J, Kittelmann S *et al* (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res* **24:** 1517-1525.

Snelling TJ, Wallace RJ (2017). The rumen microbial metaproteome as revealed by SDS-PAGE. *BMC Microbiol* **17:** 9.

Tveit AT, Urich T, Svenning MM (2014). Metatranscriptomic analysis of arctic peat soil microbiota. *Appl Environ Microbiol* **80:** 5761-5772.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3:** e2527.

Wallace RJ, Rooke JA, Duthie CA, Hyslop JJ, Ross DW, McKain N *et al* (2014). Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. *Sci Rep* **4:** 5892.

Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A *et al* (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5:** 27.

Zhou M, Hernandez-Sanabria E, Le LG (2009). Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Appl Environ Microbiol* **75:** 6524-6533.

Bibliography

Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL *et al* (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* **8:** e1002358.

Aggrey SE, Karnuah AB, Sebastian B, Anthony NB (2010). Genetic properties of feed efficiency parameters in meat-type chickens. *Genet Sel Evol* **42:** 25.

Alexandratos N, Bruinsma J (2012). World agriculture towards 2030/2050: the 2012 revision. ESA Working paper Rome, FAO.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

AlZahal O, Li F, Guan LL, Walker ND, McBride BW (2017). Factors influencing ruminal bacterial community diversity and composition and microbial fibrolytic enzyme abundance in lactating dairy cows with a focus on the role of active dry yeast. *J Dairy Sci*.

Ametaj BN, Zebeli Q, Saleem F, Psychogios N, Lewis MJ, Dunn SM *et al* (2010). Metabolomics reveals unhealthy alterations in rumen metabolism with increased proportion of cereal grain in the diet of dairy cows. *Metabolomics* **6:** 583-594.

Anders S, Pyl PT, Huber W (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:** 166-169.

Angela RM, Jean-Pierre J, John N (2000). Methane production by ruminants: its contribution to global warming. *Ann Zootech* **49:** 231-253.

Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, Parnell PF (2001). Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. *J Anim Sci* **79:** 2805-2811.

Aschenbach JR, Penner GB, Stumpff F, Gabel G (2011). Ruminant Nutrition Symposium: Role of fermentation acid absorption in the regulation of ruminal pH. *J Anim Sci* **89:** 1092-1107.

Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3:** e1029.

Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F *et al* (2017). Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *mSystems* **2**.

Avgustin G, Wallace RJ, Flint HJ (1997). Phenotypic diversity among ruminal isolates of Prevotella ruminicola: proposal of Prevotella brevis sp. nov., Prevotella bryantii sp. nov., and Prevotella albensis sp. nov. and redefinition of Prevotella ruminicola. *Int J Syst Bacteriol* **47:** 284-288.

Bach A, Calsamiglia S, Stern MD (2005). Nitrogen metabolism in the rumen. *J Dairy Sci* **88 Suppl 1:** E9-21.

Baker BJ, Sheik CS, Taylor CA, Jain S, Bhasi A, Cavalcoli JD *et al* (2013). Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling. *Isme j* **7:** 1962-1973.

Barberan A, Bates ST, Casamayor EO, Fierer N (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *Isme j* **6:** 343-351.

Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W, Thomas MB (2007). A validated whole-genome association study of efficient food conversion in cattle. *Genetics* **176:** 1893-1905.

Basarab JA, Price MA, Aalhus JL, Okine EK, Snelling WM, Lyle KL (2003). Residual feed intake and body composition in young growing cattle. *Can J Anim Sci* **83:** 189-204.

Basarab JA, Colazo MG, Ambrose DJ, Novak S, McCartney D, Baron VS (2011). Residual feed intake adjusted for backfat thickness and feeding frequency is independent of fertility in beef heifers. *Can J Anim Sci* **91:** 573-584.

Beauchemin KA, Eriksen L, Norgaard P, Rode LM (2008). Short communication: salivary secretion during meals in lactating dairy cattle. *J Dairy Sci* **91:** 2077-2081.

Belanche A, Kingston-Smith AH, Newbold CJ (2016). An Integrated Multi-Omics Approach Reveals the Effects of Supplementing Grass or Grass Hay with Vitamin E on the Rumen Microbiome and Its Function. *Front Microbiol* **7:** 905.

Benjamini Y, 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:** 289-300.

Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J *et al* (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107:** 18933-18938.

Bensoussan L, Morais S, Dassa B, Friedman N, Henrissat B, Lombard V *et al* (2017). Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome. *Environ Microbiol* **19:** 185-197.

Béra-Maillet C, Mosoni P, Kwasiborski A, Suau F, Ribot Y, 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. *J Microbiol Methods* **77:** 8-16.

Berg Miller ME, Yeoman CJ, Chia N, Tringe SG, Angly FE, Edwards RA *et al* (2012). Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environ Microbiol* **14:** 207-227.

Bergman EN (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev* **70:** 567-590.

Berry D, Schwab C, Milinovich G, Reichert J, Ben Mahfoudh K, Decker T *et al* (2012). Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis. *Isme j* **6:** 2091-2106.

Bevans DW, Beauchemin KA, Schwartzkopf-Genswein KS, McKinnon JJ, McAllister TA (2005). Effect of rapid or gradual grain adaptation on subacute acidosis and feed intake by feedlot cattle. *J Anim Sci* **83:** 1116-1132.

Biddle A, Stewart L, Blanchard J, Leschine S (2013). Untangling the genetic basis of fibrolytic specialization by lachnospiraceae and ruminococcaceae in diverse gut communities. *Diversity* **5:** 627-640.

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *Isme j* **7:** 2061-2068.

Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT *et al* (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biol* **16:** 191.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R *et al* (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* **10:** 57-59.

Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114-2120.

Brady A, Salzberg SL (2009). Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods* **6:** 673-676.

Bremer H, Dennis PP (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* **3**.

Broderick GA, Reynal SM (2009). Effect of source of rumen-degraded protein on production and ruminal metabolism in lactating dairy cows. *J Dairy Sci* **92:** 2822-2834.

Brugere JF, Borrel G, Gaci N, Tottey W, O'Toole PW, Malpuech-Brugere C (2014). Archaebiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut Microbes* **5:** 5-10.

Brulc JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC, Dinsdale EA *et al* (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci U S A* **106:** 1948-1953.

Buchfink B, Xie C, Huson DH (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12:** 59-60.

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP *et al* (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* **41:** D226-D232.

Campanaro S, Treu L, Cattani M, Kougias PG, Vendramin V, Schiavon S *et al* (2017). In vitro fermentation of key dietary compounds with rumen fluid: A genome-centric perspective. *Sci Total Environ* **584-585:** 683-691.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N *et al* (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *Isme j* **6:** 1621-1624.

Capper JL, Cady RA (2012). A comparison of the environmental impact of Jersey compared with Holstein milk for cheese production. *J Dairy Sci* **95:** 165-176.

Carberry CA, Kenny DA, Han S, McCabe MS, Waters SM (2012). Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. *Appl Environ Microbiol* **78:** 4949-4958.

Carberry CA, Waters SM, Kenny DA, Creevey CJ (2014). Rumen methanogenic genotypes differ in abundance according to host residual feed intake phenotype and diet type. *Appl Environ Microbiol* **80:** 586-594.

Carstens GE, Tedeschi LO (2016). Defining feed efficiency in beef cattle. *Proceedings of Beef Improvement Federation 38th Annual Research Symposium and Annual Meeting*; 2006; Choctaw,Mississippi.

Ceacero TM, Mercadante ME, Cyrillo JN, Canesin RC, Bonilha SF, de Albuquerque LG (2016). Phenotypic and Genetic Correlations of Feed Efficiency Traits with Growth and Carcass Traits in Nellore Cattle Selected for Postweaning Weight. *PLoS One* **11:** e0161366.

Chakravorty S, Helb D, Burday M, Connell N, Alland D (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* **69:** 330-339.

Chen Y, Penner GB, Li M, Oba M, Guan LL (2011). Changes in bacterial diversity associated with epithelial tissue in the beef cow rumen during the transition to a high-grain diet. *Appl Environ Microbiol* **77:** 5770-5781.

Chomczynski P, Sacchi N (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162:** 156-159.

Colucci PE, Chase LE, Van Soest PJ (1982). Feed Intake, Apparent Diet Digestibility, and Rate of Particulate Passage in Dairy Cattle. *J Dairy Sc* **65:** 1445-1456.

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. *Front Microbiol* **8:** 67.

Cox MP, Peterson DA, Biggs PJ (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11:** 485.

Creevey CJ, Kelly WJ, Henderson G, Leahy SC (2014). Determining the culturability of the rumen bacterial microbiome. *Microb Biotechnol* **7:** 467-479.

Crews DH, Jr. (2005). Genetics of efficient feed utilization and national cattle evaluation: a review. *Genet Mol Res* **4:** 152-165.

Dai X, Zhu Y, Luo Y, Song L, Liu D, Liu L *et al* (2012). Metagenomic insights into the fibrolytic microbiome in yak rumen. *PLoS One* **7:** e40430.

Danielsson R, Dicksved J, Sun L, Gonda H, Muller B, Schnurer A *et al* (2017). Methane Production in Dairy Cows Correlates with Rumen Methanogenic and Bacterial Community Structure. *Front Microbiol* **8:** 226.

Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y (2015). Genome-Wide Association Studies of the Human Gut Microbiota. *PLoS One* **10:** e0140301.

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE *et al* (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505:** 559-563.

de Menezes A, Clipson N, Doyle E (2012). Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. *Environ Microbiol* **14:** 2577-2588.

de Oliveira PS, Cesar AS, do Nascimento ML, Chaves AS, Tizioto PC, Tullio RR *et al* (2014). Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genet* **15:** 100.

Deng L, He C, Zhou Y, Xu L, Xiong H (2017). Ground transport stress affects bacteria in the rumen of beef cattle: A real-time PCR analysis. *Anim Sci J* **88:** 790-797.

Deng W, Xi D, Mao H, Wanapat M (2008). The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. *Mol Biol Rep* **35:** 265-274.

Denman SE, Martinez Fernandez G, Shinkai T, Mitsumori M, McSweeney CS (2015). Metagenomic analysis of the rumen microbial community following inhibition of methane formation by a halogenated methane analog. *Front Microbiol* **6:** 1087.

Derakhshani H, Tun HM, Cardoso FC, Plaizier JC, Khafipour E, Loor JJ (2016). Linking Peripartal Dynamics of Ruminal Microbiota to Dietary Changes and Production Parameters. *Front Microbiol* **7:** 2143.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069-5072.

Driscoll CB, Otten TG, Brown NM, Dreher TW (2017). Towards long-read metagenomics: complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture. *Stand Genomic Sci* **12:** 9.

Dubin K, Callahan MK, Ren B, Khanin R, Viale A, Ling L *et al* (2016). Intestinal microbiome analyses identify melanoma patients at risk for checkpoint-blockade-induced colitis. *Nat Commun* **7:** 10391.

Durunna ON, Mujibi FD, Goonewardene L, Okine EK, Basarab JA, Wang Z *et al* (2011). Feed efficiency differences and reranking in beef steers fed grower and finisher diets. *J Anim Sci* **89:** 158-167.

Eddy SR (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol* **7:** e1002195.

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194-2200.

Edwards JE, McEwan NR, Travis AJ, Wallace RJ (2004). 16S rDNA library-based analysis of ruminal bacterial diversity. *Antonie van Leeuwenhoek* **86:** 263-281.

Eisler MC, Lee MR, Tarlton JF, Martin GB, Beddington J, Dungait JA *et al* (2014). Agriculture: Steps to sustainable livestock. *Nature* **507:** 32-34.

Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ *et al* (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145:** 745-757.

Ellison MJ, Conant GC, Cockrum RR, Austin KJ, Truong H, Becchi M *et al* (2014). Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. *DNA Res* **21:** 115-125.

Embley TM, Horner DA, Hirt RP (1997). Anaerobic eukaryote evolution: hydrogenosomes as biochemically modified mitochondria? *Trends Ecol Evol* **12:** 437-441.

Endelman JB (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome* **4:** 250-255.

Eren AM, Sogin ML, Morrison HG, Vineis JH, Fisher JC, Newton RJ *et al* (2015). A single genus in the gut microbiome reflects host preference and specificity. *Isme j* **9:** 90-100.

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J *et al* (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* **8:** e1002606.

Findley SD, Mormile MR, Sommer-Hurley A, Zhang XC, Tipton P, Arnett K *et al* (2011). Activity-based metagenomic screening and biochemical characterization of bovine ruminal protozoan glycoside hydrolases. *Appl Environ Microbiol* **77:** 8106-8113.

Firkins JL (1996). Maximizing microbial protein synthesis in the rumen. *J Nutr* **126:** 1347s-1354s.

Firkins JL, 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:** 1450-1470.

Flint HJ, Bayer EA (2008). Plant cell wall breakdown by anaerobic microorganisms from the Mammalian digestive tract. *Ann N Y Acad Sci* **1125:** 280-288.

Fox GE, Wisotzkey JD, Jurtshuk P, Jr. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42:** 166-170.

France J, Theodorou MK, Davies D (1990). Use of zoospore concentrations and life cycle parameters in determining the population of anaerobic fungi in the rumen ecosystem. *J Theor Biol* **147:** 413-422.

Frank DN, Robertson CE, Hamm CM, Kpadeh Z, Zhang T, Chen H *et al* (2011). Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* **17:** 179-184.

Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM *et al* (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* **111:** E2329-2338.

Friedman J, Alm EJ (2012). Inferring correlation networks from genomic survey data. *PLoS Comput Biol* **8:** e1002687.

Gaci N, Borrel G, Tottey W, O'Toole PW, Brugere JF (2014). Archaea and the human gut: new beginning of an old story. *World J Gastroenterol* **20:** 16062-16078.

Gaidos E, Rusch A, Ilardo M (2011). Ribosomal tag pyrosequencing of DNA and RNA from benthic coral reef microbiota: community spatial structure, rare members and nitrogen-cycling guilds. *Environ Microbiol* **13:** 1138-1152.

Garcia-Vallve S, Romeu A, Palau J (2000). Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Mol Biol Evol* **17:** 352-361.

Ghali I, Shinkai T, Mitsumori M (2016). Mining of luxS genes from rumen microbial consortia by metagenomic and metatranscriptomic approaches. *Anim Sci J* **87:** 666-673.

Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ *et al* (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol* **13:** R23.

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R, Butler D *et al* (2014). *ASReml user guide. Release 4.1 structural specification*: VSN International Ltd, Hemel Hempstead, UK.

Golder HM, Denman SE, McSweeney C, Celi P, Lean IJ (2014). Ruminal bacterial community shifts in grain-, sugar-, and histidine-challenged dairy heifers. *J Dairy Sci* **97:** 5131-5150.

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R *et al* (2014). Human genetics shape the gut microbiome. *Cell* **159:** 789-799.

Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C *et al* (2016). Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe* **19:** 731-743.

Goopy JP, Donaldson A, Hegarty R, Vercoe PE, Haynes F, Barnett M *et al* (2014). Low-methane yield sheep have smaller rumens and shorter rumen retention time. *Br J Nutr* **111:** 578-585.

Gordon GL, Phillips MW (1998). The role of anaerobic gut fungi in ruminants. *Nutr Res Rev* **11:** 133-168.

Granja-Salcedo YT, Ramirez-Uscategui RA, Machado EG, Duarte Messana J, Takeshi Kishi L, Lino Dias AV *et al* (2017). Studies on bacterial community composition are affected by the time and storage method of the rumen content. *PLoS One* **12:** e0176701.

Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ (2000). Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol* **66:** 5488-5491.

Gruninger RJ, Gong X, Forster RJ, McAllister TA (2014a). Biochemical and kinetic characterization of the multifunctional beta-glucosidase/beta-xylosidase/alpha-arabinosidase, Bgxa1. *Appl Microbiol Biotechnol* **98:** 3003-3012.

Gruninger RJ, Puniya AK, Callaghan TM, Edwards JE, Youssef N, Dagar SS *et al* (2014b). Anaerobic fungi (phylum Neocallimastigomycota): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiol Ecol* **90:** 1-17.

Guan LL, Nkrumah JD, Basarab JA, Moore SS (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol Lett* **288:** 85-91.

Guyader J, Eugene M, Noziere P, Morgavi DP, Doreau M, Martin C (2014). Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. *Animal* **8:** 1816-1825.

Haange SB, Jehmlich N (2016). Proteomic interrogation of the gut microbiota: potential clinical impact. *Expert Rev Proteomics* **13:** 535-537.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G *et al* (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21:** 494-504.

Hargrove JL, Schmidt FH (1989). The role of mRNA and protein stability in gene expression. *Faseb j* **3:** 2360-2370.

Hedderich R, Whitman WB (2006). Physiology and biochemistry of the methane-producing Archaea. *The prokaryotes*. Springer. pp 1050-1079.

Hegarty RS, Goopy JP, Herd RM, McCorkell B (2007). Cattle selected for lower residual feed intake have reduced daily methane production. *J Anim Sci* **85:** 1479-1486.

Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, Noel SJ *et al* (2013). Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One* **8:** e74787.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH (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.

Herd RM, Oddy VH, Richardson EC (2004). Biological basis for variation in residual feed intake in beef cattle. 1. Review of potential mechanisms. *Australian Journal of Experimental Agriculture* **44:** 423-430.

Herd RM, Arthur PF (2009). Physiological basis for residual feed intake. *J Anim Sci* **87:** E64-71.

Hernandez J, Benedito JL, Abuelo A, Castillo C (2014). Ruminal acidosis in feedlot: from aetiology to prevention. *ScientificWorldJournal* **2014:** 702572.

Hernandez-Sanabria E, Guan LL, Goonewardene LA, Li M, Mujibi DF, Stothard P *et al* (2010). Correlation of particular bacterial PCR-denaturing gradient gel electrophoresis patterns

with bovine ruminal fermentation parameters and feed efficiency traits. *Appl Environ Microbiol* **76:** 6338-6350.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Durunna ON, Moore SS, Guan LL (2012). Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. *Appl Environ Microbiol* **78:** 1203-1214.

Hernandez-Sanabria E, Goonewardene LA, Wang Z, Zhou M, Moore SS, Guan LL (2013). Influence of sire breed on the interplay among rumen microbial populations inhabiting the rumen liquid of the progeny in beef cattle. *PLoS One* **8:** e58461.

Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G *et al* (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331:** 463-467.

Holmes E, Li JV, Marchesi JR, Nicholson JK (2012). Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab* **16:** 559-564.

Hong S, Bunge J, Leslin C, Jeon S, Epstein SS (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *Isme j* **3:** 1365-1373.

Hoque MA, Arthur PF, Hiramoto K, Oikawa T (2006). Genetic relationship between different measures of feed efficiency and its component traits in Japanese Black (Wagyu) bulls. *Livestock Science* **99:** 111-118.

Huber JA, Morrison HG, Huse SM, Neal PR, Sogin ML, Mark Welch DB (2009). Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ Microbiol* **11:** 1292-1302.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC *et al* (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44:** D286-293.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486:** 207-214.

Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S *et al* (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol* **12:** e1004957.

Huws SA, Kim EJ, Lee MR, Scott MB, Tweed JK, Pinloche E *et al* (2011). As yet uncultured bacteria phylogenetically classified as Prevotella, Lachnospiraceae incertae sedis and unclassified Bacteroidales, Clostridiales and Ruminococcaceae may play a predominant role in ruminal biohydrogenation. *Environ Microbiol* **13:** 1500-1512.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11:** 119.

İnceoğlu Ö, Llirós M, Crowe SA, García-Armisen T, Morana C, Darchambeau F *et al* (2015). Vertical Distribution of Functional Potential and Active Microbial Communities in Meromictic Lake Kivu. *Microb Ecol*.

Ishaq SL, Sundset MA, Crouse J, Wright AG (2015). High-throughput DNA sequencing of the moose rumen from different geographical locations reveals a core ruminal methanogenic archaeal diversity and a differential ciliate protozoal diversity. *Microb Genom* **1:** e000034.

Jami E, Mizrahi I (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS One* **7:** e33306.

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

Jami E, White BA, Mizrahi I (2014). Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *PLoS One* **9:** e85423.

Janssen PH, Kirs M (2008). Structure of the archaeal community of the rumen. *Appl Environ Microbiol* **74:** 3619-3625.

Janssen PH (2010). Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. *Animal Feed Science and Technology* **160:** 1-22.

Jewell KA, McCormick CA, Odt CL, Weimer PJ, Suen G (2015). Ruminal Bacterial Community Composition in Dairy Cows Is Dynamic over the Course of Two Lactations and Correlates with Feed Efficiency. *Appl Environ Microbiol* **81:** 4697-4710.

Jeyanathan J, Kirs M, Ronimus RS, Hoskin SO, Janssen PH (2011). Methanogen community structure in the rumens of farmed sheep, cattle and red deer fed different diets. *FEMS Microbiol Ecol* **76:** 311-326.

Johnson KA, Johnson DE (1995). Methane emissions from cattle. *J Anim Sci* **73:** 2483-2492.

Jose VL, More RP, Appoothy T, Arun AS (2017). In depth analysis of rumen microbial and carbohydrate-active enzymes profile in Indian crossbred cattle. *Syst Appl Microbiol* **40:** 160-170.

Jost L (2007). Partitioning diversity into independent alpha and beta components. *Ecology* **88:** 2427-2439.

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:** 56.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40:** D109-114.

Kang S, Denman SE, Morrison M, Yu Z, McSweeney CS (2009). An efficient RNA extraction method for estimating gut microbial diversity by polymerase chain reaction. *Curr Microbiol* **58:** 464-471.

Kang SH, Evans P, Morrison M, McSweeney C (2013). Identification of metabolically active proteobacterial and archaeal communities in the rumen by DNA- and RNA-derived 16S rRNA gene. *J Appl Microbiol* **115:** 644-653.

Karisa B, Moore S, Plastow G (2014). Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle. *Anim Sci J* **85:** 374-387.

Kellermayer R, Dowd SE, Harris RA, Balasa A, Schaible TD, Wolcott RD *et al* (2011). Colonic mucosal DNA methylation, immune response, and microbiome patterns in Toll-like receptor 2-knockout mice. *Faseb j* **25:** 1449-1460.

Khachatryan ZA, Ktsoyan ZA, Manukyan GP, Kelly D, Ghazaryan KA, Aminov RI (2008). Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS One* **3:** e3064.

Khafipour E, Li S, Plaizier JC, Krause DO (2009). Rumen microbiome composition determined using two nutritional models of subacute ruminal acidosis. *Appl Environ Microbiol* **75:** 7115-7124.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14:** R36.

Kim M, Morrison M, Yu Z (2011). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol Ecol* **76:** 49-63.

Kim M, Yu Z (2014). Variations in 16S rRNA-based microbiome profiling between pyrosequencing runs and between pyrosequencing facilities. *J Microbiol* **52:** 355-365.

Kittelmann S, Naylor GE, Koolaard JP, Janssen PH (2012). A proposed taxonomy of anaerobic fungi (class neocallimastigomycetes) suitable for large-scale sequence-based community structure analysis. *PLoS One* **7:** e36866.

Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI *et al* (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS One* **8:** e47879.

Kittelmann S, Pinares-Patino CS, Seedorf H, Kirk MR, Ganesh S, McEwan JC *et al* (2014). Two different bacterial community types are linked with the low-methane emission trait in sheep. *PLoS One* **9:** e103171.

Klieve AV, Swain RA (1993). Estimation of ruminal bacteriophage numbers by pulsed-field gel electrophoresis and laser densitometry. *Appl Environ Microbiol* **59:** 2299-2303.

Klieve AV, O'Leary MN, McMillen L, Ouwerkerk D (2007). Ruminococcus bromii, identification and isolation as a dominant community member in the rumen of cattle fed a barley diet. *J Appl Microbiol* **103:** 2065-2073.

Klieve AV, McLennan SR, Ouwerkerk D (2012). Persistence of orally administered Megasphaera elsdenii and Ruminococcus bromii in the rumen of beef cattle fed a high grain (barley) diet. *Anim Prod Sci* **52:** 297-304.

Koch RM, Swiger LA, Chambers D, Gregory KE (1963). Efficiency of Feed Use in Beef Cattle. *J Anim Sci* **22:** 486-494.

Koeppel AF, Wu M (2013). Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. *Nucleic Acids Res* **41:** 5175-5188.

Koetschan C, Kittelmann S, Lu J, Al-Halbouni D, Jarvis GN, Muller T *et al* (2014). Internal transcribed spacer 1 secondary structure analysis reveals a common core throughout the anaerobic fungi (Neocallimastigomycota). *PLoS One* **9:** e91928.

Koike S, Pan J, Kobayashi Y, Tanaka K (2003). Kinetics of In Sacco Fiber-Attachment of Representative Ruminal Cellulolytic Bacteria Monitored by Competitive PCR. *Journal of Dairy Science* **86:** 1429-1435.

Kong RS, Liang G, Chen Y, Stothard P, Guan le L (2016). Transcriptome profiling of the rumen epithelium of beef cattle differing in residual feed intake. *BMC Genomics* **17:** 592.

Kopylova E, Noe L, Touzet H (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28:** 3211-3217.

Korpela K, Salonen A, Virta LJ, Kekkonen RA, Forslund K, Bork P *et al* (2016). Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nat Commun* **7:** 10410.

Kostic AD, Howitt MR, Garrett WS (2013). Exploring host-microbiota interactions in animal models and humans. *Genes Dev* **27:** 701-718.

Kozich J, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl Environ Microbiol* **79:** 5112-5120.

La Reau AJ, Meier-Kolthoff JP, Suen G (2016). Sequence-based analysis of the genus Ruminococcus resolves its phylogeny and reveals strong host association. *Microb Genom* **2:** e000099.

Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P *et al* (2016). Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol* **1:** 16203.

Lamb GC, Maddock T (2009). Feed efficiency in cows. *Cash, Cows, and Calves* **1:** 35-42.

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357-359.

Lanzen A, Jorgensen SL, Bengtsson MM, Jonassen I, Ovreas L, Urich T (2011). Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiol Ecol* **77:** 577-589.

Lawrence P, Kenny DA, Earley B, Crews DH, Jr., McGee M (2011). Grass silage intake, rumen and blood variables, ultrasonic and body measurements, feeding behavior, and activity in pregnant beef heifers differing in phenotypic residual feed intake. *J Anim Sci* **89:** 3248-3261.

Leamy LJ, Kelly SA, Nietfeldt J, Legge RM, Ma F, Hua K *et al* (2014). Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biol* **15:** 552.

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW *et al* (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442:** 806-809.

Li C, Basarab J, Snelling WM, Benkel B, Murdoch B, Moore SS (2002). The identification of common haplotypes on bovine chromosome 5 within commercial lines of Bos taurus and their associations with growth traits. *J Anim Sci* **80:** 1187-1194.

Li D, Liu CM, Luo R, Sadakane K, Lam TW (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31:** 1674-1676.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al* (2009a). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078-2079.

Li H, Durbin R (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26:** 589-595.

Li M, Penner GB, Hernandez-Sanabria E, Oba M, Guan LL (2009b). Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen. *J Appl Microbiol* **107:** 1924-1934.

Li RW, Giarrizzo JG, Wu S, Li W, Duringer JM, Craig AM (2014). Metagenomic insights into the RDX-degrading potential of the ovine rumen microbiome. *PLoS One* **9:** e110505.

Li T, Chiang JY (2015). Bile acids as metabolic regulators. *Curr Opin Gastroenterol* **31:** 159-165.

Li Y, Leahy SC, Jeyanathan J, Henderson G, Cox F, Altermann E *et al* (2016). The complete genome sequence of the methanogenic archaeon ISO4-H5 provides insights into the methylotrophic lifestyle of a ruminal representative of the Methanomassiliicoccales. *Stand Genomic Sci* **11:** 59.

Lim S, Seo J, Choi H, Yoon D, Nam J, Kim H *et al* (2013). Metagenome Analysis of Protein Domain Collocation within Cellulase Genes of Goat Rumen Microbes. *Asian-Australas J Anim Sci* **26:** 1144-1151.

Liu L, Li Y, Li S, Hu N, He Y, Pong R *et al* (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012:** 251364.

Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento H *et al* (2014). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* **16:** 2659-2671.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42:** D490-495.

Lopes LD, de Souza Lima AO, Taketani RG, Darias P, da Silva LR, Romagnoli EM *et al* (2015). Exploring the sheep rumen microbiome for carbohydrate-active enzymes. *Antonie Van Leeuwenhoek* **108:** 15-30.

Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J *et al* (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1:** 18.

Ma J, Coarfa C, Qin X, Bonnen PE, Milosavljevic A, Versalovic J *et al* (2014). mtDNA haplogroup and single nucleotide polymorphisms structure human microbiome communities. *BMC Genomics* **15:** 257.

Mandal S, Van Treuren W, White RA, Eggesbo M, Knight R, Peddada SD (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis* **26:** 27663.

Mao F, Chen L, Vinsky M, Okine E, Wang Z, Basarab J *et al* (2013). Phenotypic and genetic relationships of feed efficiency with growth performance, ultrasound, and carcass merit traits in Angus and Charolais steers. *J Anim Sci* **91:** 2067-2076.

Marchesi JR, Adams DH, Fava F, Hermes GD, Hirschfield GM, Hold G *et al* (2015). The gut microbiota and host health: a new clinical frontier. *Gut*.

Martinez A, Ventouras LA, Wilson ST, Karl DM, Delong EF (2013). Metatranscriptomic and functional metagenomic analysis of methylphosphonate utilization by marine bacteria. *Front Microbiol* **4:** 340.

Martinez ME, Ranilla MJ, Ramos S, Tejido ML, Carro MD (2009). Effects of dilution rate and retention time of concentrate on efficiency of microbial growth, methane production, and ruminal fermentation in Rusitec fermenters. *J Dairy Sci* **92:** 3930-3938.

Martinez-Fernandez G, Denman SE, Yang C, Cheung J, Mitsumori M, McSweeney CS (2016). Methane Inhibition Alters the Microbial Community, Hydrogen Flow, and Fermentation Response in the Rumen of Cattle. *Front Microbiol* **7:** 1122.

Mason OU, Scott NM (2014). Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill **8:** 1464-1475.

Maurice CF, Haiser HJ, Turnbaugh PJ (2013). Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* **152:** 39-50.

McAllister TA, Bae HD, Jones GA, Cheng KJ (1994). Microbial attachment and feed digestion in the rumen. *Journal of Animal Science* **72:** 3004-3018.

McCann JC, Luan S, Cardoso FC, Derakhshani H, Khafipour E, Loor JJ (2016). Induction of Subacute Ruminal Acidosis Affects the Ruminal Microbiome and Epithelium. *Front Microbiol* **7:** 701.

McCuddin ZP, Carlson SA, Rasmussen MA, Franklin SK (2006). Klebsiella to Salmonella gene transfer within rumen protozoa: implications for antibiotic resistance and rumen defaunation. *Vet Microbiol* **114:** 275-284.

McDonagh MB, Herd RM, Richardson EC, Oddy VH, Archer JA, Arthur PF (2001). Meat quality and the calpain system of feedlot steers following a single generation of divergent selection for residual feed intake. *Australian Journal of Experimental Agriculture* **41:** 1013-1021.

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A *et al* (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *Isme j* **6:** 610-618.

McEwan NR, Abecia L, Regensbogenova M, Adam CL, Findlay PA, Newbold CJ (2005). Rumen microbial population dynamics in response to photoperiod. *Lett Appl Microbiol* **41:** 97-101.

McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL *et al* (2011). The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci Transl Med* **3:** 106ra106.

Medlin L, Simon N (1998). Phylogenetic Analysis of Marine Phytoplankton. In: Cooksey K (ed). *Molecular Approaches to the Study of the Ocean*. Springer Netherlands. pp 161-186.

Meehan CJ, Beiko RG (2014). A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol Evol* **6:** 703-713.

Mendes LW, Kuramae EE, Navarrete AA, van Veen JA, Tsai SM (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. *Isme j* **8:** 1577-1587.

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

Miller TL, Wolin MJ, Zhao HX, Bryant MP (1986). Characteristics of methanogens isolated from bovine rumen. *Appl Environ Microbiol* **51:** 201-202.

Million M, Lagier JC, Yahav D, Paul M (2013). Gut bacterial microbiota and obesity. *Clin Microbiol Infect* **19:** 305-313.

Minato H, Mitsumori M, Cheng KJ (1993). Attachment of microorganisms to solid substrates in the rumen. *Genetics, Biochemistry and Ecology of Lignocellulose Degradation (Shimada, K, Ohmiya, K, Kobayashi, Y, Hoshino, S, Sakka, K and Karita, S, Eds)***:** 139-145.

Moore SS, Mujibi FD, Sherman EL (2009). Molecular basis for residual feed intake in beef cattle. *J Anim Sci* **87:** E41-47.

Morgavi DP, Kelly WJ, Janssen PH, Attwood GT (2013). Rumen microbial (meta)genomics and its application to ruminant production. *Animal* **7 Suppl 1:** 184-201.

Mottawea W, Chiang CK, Muhlbauer M, Starr AE, Butcher J, Abujamel T *et al* (2016). Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. *Nat Commun* **7:** 13419.

Muhling M, Woolven-Allen J, Murrell JC, Joint I (2008). Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *Isme j* **2:** 379-392.

Mujibi FD, Moore SS, Nkrumah DJ, Wang Z, Basarab JA (2010). Season of testing and its effect on feed intake and efficiency in growing beef cattle. *J Anim Sci* **88:** 3789-3799.

Myer PR, Smith TP, Wells JE, Kuehn LA, Freetly HC (2015a). Rumen microbiome from steers differing in feed efficiency. *PLoS One* **10:** e0129174.

Myer PR, Wells JE, Smith TP, Kuehn LA, Freetly HC (2015b). Microbial community profiles of the colon from steers differing in feed efficiency. *Springerplus* **4:** 454.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* **40:** e155.

Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z *et al* (2013). Advancing Our Understanding of the Human Microbiome Using QIIME. *Methods in Enzymology*. pp 371-444.

Nayfach S, Bradley PH, Wyman SK, Laurent TJ, Williams A, Eisen JA *et al* (2015). Automated and Accurate Estimation of Gene Family Abundance from Shotgun Metagenomes. *PLoS Comput Biol* **11:** e1004573.

Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J (2014). Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS One* **9:** e94249.

Newbold CJ, de la Fuente G, Belanche A, Ramos-Morales E, McEwan NR (2015). The Role of Ciliate Protozoa in the Rumen. *Front Microbiol* **6:** 1313.

Nicolazzi EL, Caprera A, Nazzicari N, Cozzi P, Strozzi F, Lawley C *et al* (2015). SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics* **16:** 283.

Nkrumah JD, Okine EK, Mathison GW, Schmid K, Li C, Basarab JA *et al* (2006). Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle. *J Anim Sci* **84:** 145-153.

Nkrumah JD, Basarab JA, Wang Z, Li C, Price MA, Okine EK *et al* (2007a). Genetic and phenotypic relationships of feed intake and measures of efficiency with growth and carcass merit of beef cattle. *J Anim Sci* **85:** 2711-2720.

Nkrumah JD, Crews DH, Jr., Basarab JA, Price MA, Okine EK, Wang Z *et al* (2007b). Genetic and phenotypic relationships of feeding behavior and temperament with performance, feed efficiency, ultrasound, and carcass merit of beef cattle. *J Anim Sci* **85:** 2382-2390.

Nkrumah JD, Sherman EL, Li C, Marques E, Crews DH, Jr., Bartusiak R *et al* (2007c). Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. *J Anim Sci* **85:** 3170-3181.

Noel SJ, Hojberg O, Urich T, Poulsen M (2016). Draft Genome Sequence of "Candidatus Methanomethylophilus" sp. 1R26, Enriched from Bovine Rumen, a Methanogenic Archaeon Belonging to the Methanomassiliicoccales Order. *Genome Announc* **4**.

Noel SJ, Attwood GT, Rakonjac J, Moon CD, Waghorn GC, Janssen PH (2017). Seasonal changes in the digesta-adherent rumen bacterial communities of dairy cattle grazing pasture. *PLoS One* **12:** e0173819.

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27:** 824-834.

Ohene-Adjei S, Teather RM, Ivan M, Forster RJ (2007). Postinoculation protozoan establishment and association patterns of methanogenic archaea in the ovine rumen. *Appl Environ Microbiol* **73:** 4609-4618.

Ojeda P, Bobe A, Dolan K, Leone V, Martinez K (2015). Nutritional modulation of gut microbiota - the impact on metabolic disease pathophysiology. *J Nutr Biochem*.

Oki K, Toyama M, Banno T, Chonan O, Benno Y, Watanabe K (2016). Comprehensive analysis of the fecal microbiota of healthy Japanese adults reveals a new bacterial lineage associated with a phenotype characterized by a high frequency of bowel movements and a lean body type. *BMC Microbiol* **16:** 284.

Olfert ED, Cross BM, McWilliams AA (1993). Guide to the care and use of experimental steers. *Canadian Council on Animal Care, Ottawa, ON, Canada*.

Ondov BD, Bergman NH, Phillippy AM (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12:** 385.

Org E, Mehrabian M, Parks BW, Shipkova P, Liu X, Drake TA *et al* (2016). Sex differences and hormonal effects on gut microbiota composition in mice. *Gut Microbes* **7:** 313-322.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M *et al* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33:** 5691-5702.

Palomo A, Jane Fowler S, Gulay A, Rasmussen S, Sicheritz-Ponten T, Smets BF (2016). Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp. *Isme j* **10:** 2569-2581.

Pandey RV, Pabinger S, Kriegner A, Weinhausel A (2016). ClinQC: a tool for quality control and cleaning of Sanger and NGS data in clinical research. *BMC Bioinformatics* **17:** 56.

Patel RK, Jain M (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **7:** e30619.

Pathak AK (2008). Various factors affecting microbial protein synthesis in the rumen. *Vet World* **1:** 186-189.

Patience JF, Rossoni-Serao MC, Gutierrez NA (2015). A review of feed efficiency in swine: biology and application. *J Anim Sci Biotechnol* **6:** 33.

Paz HA, Anderson CL, Muller MJ, Kononoff PJ, Fernando SC (2016). Rumen Bacterial Community Composition in Holstein and Jersey Cows Is Different under Same Dietary Condition and Is Not Affected by Sampling Method. *Front Microbiol* **7:** 1206.

Perkins SD, Key CN, Garrett CF, Foradori CD, Bratcher CL, Kriese-Anderson LA *et al* (2014). Residual feed intake studies in Angus-sired cattle reveal a potential role for hypothalamic gene expression in regulating feed efficiency. *J Anim Sci* **92:** 549-560.

Petri RM, Schwaiger T, Penner GB, Beauchemin KA, Forster RJ, McKinnon JJ *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:** e83424.

Pitta DW, Pinchak WE, Indugu N, Vecchiarelli B, Sinha R, Fulford JD (2016). Metagenomic Analysis of the Rumen Microbiome of Steers with Wheat-Induced Frothy Bloat. *Front Microbiol* **7:** 689.

Plaizier JC, Li S, Danscher AM, Derakshani H, Andersen PH, Khafipour E (2017). Changes in Microbiota in Rumen Digesta and Feces Due to a Grain-Based Subacute Ruminal Acidosis (SARA) Challenge. *Microb Ecol*.

Pope PB, Smith W, Denman SE, Tringe SG, Barry K, Hugenholtz P *et al* (2011). Isolation of Succinivibrionaceae implicated in low methane emissions from Tammar wallabies. *Science* **333:** 646-648.

Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC *et al* (2012). Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PLoS One* **7:** e38571.

Poulsen LK, Ballard G, Stahl DA (1993). Use of rRNA fluorescence in situ hybridization for measuring the activity of single cells in young and established biofilms. *Appl Environ Microbiol* **59:** 1354-1360.

Poulsen M, Schwab C, Jensen BB, Engberg RM, Spang A, Canibe N *et al* (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat Commun* **4:** 1428.

Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP *et al* (2007). The role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5:** 384-392.

Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* **40:** D130-135.

Pundir S, Martin MJ, O'Donovan C (2017). UniProt Protein Knowledgebase. *Methods Mol Biol* **1558:** 41-55.

Purushe J, Fouts DE, Morrison M, White BA, Mackie RI, Coutinho PM *et al* (2010). Comparative genome analysis of Prevotella ruminicola and Prevotella bryantii: insights into their environmental niche. *Microb Ecol* **60:** 721-729.

Qi M, Wang P, O'Toole N, Barboza PS, Ungerfeld E, Leigh MB *et al* (2011). Snapshot of the eukaryotic gene expression in muskoxen rumen-A metatranscriptomic approach. *PLoS ONE* **6**.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C *et al* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464:** 59-65.

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 Res* **41:** D590-D596.

R Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria.

Racca AW, Beck AE, McMillin MJ, Korte FS, Bamshad MJ, Regnier M (2015). The embryonic myosin R672C mutation that underlies Freeman-Sheldon syndrome impairs cross-bridge detachment and cycling in adult skeletal muscle. *Hum Mol Genet* **24:** 3348-3358.

Ramayo-Caldas Y, Mach N, Lepage P, Levenez F, Denis C, Lemonnier G *et al* (2016). Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *Isme j* **10:** 2973-2977.

Rauw WM, Soler J, Tibau J, Reixach J, Gomez Raya L (2006). The relationship between residual feed intake and feed intake behavior in group-housed Duroc barrows. *J Anim Sci* **84:** 956-962.

Rey FE, Faith JJ, Bain J, Muehlbauer MJ, Stevens RD, Newgard CB *et al* (2010). Dissecting the in vivo metabolic potential of two human gut acetogens. *J Biol Chem* **285:** 22082-22090.

Rho M, Tang H, Ye Y (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* **38:** e191.

Ricard G, McEwan NR, Dutilh BE, Jouany JP, 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.

Richardson EC, Herd RM, Oddy VH, Thompson JM, Archer JA, Arthur PF (2001). Body composition and implications for heat production of Angus steer progeny of parents selected for and against residual feed intake. *Australian Journal of Experimental Agriculture* **41:** 1065-1072.

Richardson EC, Herd RM (2004). Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. *Australian Journal of Experimental Agriculture* **44:** 431-440.

Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM *et al* (2014). Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* **2:** e545.

Rius AG, Kittelmann S, Macdonald KA, Waghorn GC, Janssen PH, Sikkema E (2012). Nitrogen metabolism and rumen microbial enumeration in lactating cows with divergent residual feed intake fed high-digestibility pasture. *J Dairy Sci* **95:** 5024-5034.

Robinson MD, Smyth GK (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23:** 2881-2887.

Roehe R, Dewhurst RJ, Duthie CA, Rooke JA, McKain N, Ross DW *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 Genet* **12:** e1005846.

Rohwer F, Thurber RV (2009). Viruses manipulate the marine environment. *Nature* **459:** 207-212.

Rolf MM, Taylor JF, Schnabel RD, McKay SD, McClure MC, Northcutt SL *et al* (2012). Genome-wide association analysis for feed efficiency in Angus cattle. *Anim Genet* **43:** 367-374.

Rooks MG, Veiga P, Wardwell-Scott LH, Tickle T, Segata N, Michaud M *et al* (2014). Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission **8:** 1403-1417.

Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM *et al* (2012). High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* **13:** 53.

Ross EM, Moate PJ, Marett L, Cocks BG, Hayes BJ (2013a). Investigating the effect of two methane-mitigating diets on the rumen microbiome using massively parallel sequencing. *J Dairy Sci* **96:** 6030-6046.

Ross EM, Moate PJ, Marett LC, Cocks BG, Hayes BJ (2013b). Metagenomic predictions: from microbiome to complex health and environmental phenotypes in humans and cattle. *PLoS One* **8:** e73056.

Ross EM, Petrovski S, Moate PJ, Hayes BJ (2013c). Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. *BMC Microbiol* **13:** 242.

Rossi-Tamisier M, Benamar S, Raoult D, Fournier PE (2015). Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. *Int J Syst Evol Microbiol* **65:** 1929-1934.

Russell JB, Rychlik JL (2001). Factors that alter rumen microbial ecology. *Science* **292:** 1119-1122.

Saleem F, Ametaj BN, Bouatra S, Mandal R, Zebeli Q, Dunn SM *et al* (2012). A metabolomics approach to uncover the effects of grain diets on rumen health in dairy cows. *J Dairy Sci* **95:** 6606-6623.

Sanders JG, Beichman AC, Roman J, Scott JJ, Emerson D, McCarthy JJ *et al* (2015). Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. *Nat Commun* **6:** 8285.

Santana MHA, Rossi P, Almeida R, Cucco DC (2012). Feed efficiency and its correlations with carcass traits measured by ultrasound in Nellore bulls. *Livestock Science* **145:** 252-257.

Santschi DE, Berthiaume R, Matte JJ, Mustafa AF, Girard CL (2005). Fate of supplementary B-vitamins in the gastrointestinal tract of dairy cows. *J Dairy Sci* **88:** 2043-2054.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K *et al* (2011). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **39:** D38-51.

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. *J Dairy Sci* **100:** 2765-2783.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537-7541.

Schmieder R, Edwards R (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27:** 863-864.

Schwab C, Berry D, Rauch I, Rennisch I, Ramesmayer J, Hainzl E *et al* (2014). Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *Isme j* **8:** 1101-1114.

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J *et al* (2011). Global quantification of mammalian gene expression control. *Nature* **473:** 337-342.

Schwieger F, Tebbe CC (1998). A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. *Appl Environ Microbiol* **64:** 4870-4876.

Seedorf H, Kittelmann S, Henderson G, Janssen PH (2014). RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. *PeerJ* **2:** e494.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS *et al* (2011). Metagenomic biomarker discovery and explanation. *Genome Biol* **12:** R60.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9:** 811-814.

Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C (2013). Computational meta'omics for microbial community studies. *Mol Syst Biol* **9:** 666.

Shabat SK, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Berg Miller ME *et al* (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *Isme j*.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498-2504.

Shen JS, Chai Z, Song LJ, Liu JX, Wu YM (2012). Insertion depth of oral stomach tubes may affect the fermentation parameters of ruminal fluid collected in dairy cows. *J Dairy Sci* **95:** 5978-5984.

Sherman EL, Nkrumah JD, Murdoch BM, Li C, Wang Z, Fu A *et al* (2008). Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. *J Anim Sci* **86:** 1-16.

Sherman EL, Nkrumah JD, Li C, Bartusiak R, Murdoch B, Moore SS (2009). Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. *J Anim Sci* **87:** 37-45.

Shi W, Moon CD, Leahy SC, Kang D, Froula J, Kittelmann S *et al* (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res* **24:** 1517-1525.

Shriver BJ, Hoover WH, Sargent JP, Crawford RJ, Jr., Thayne WV (1986). Fermentation of a High Concentrate Diet as Affected by Ruminal pH and Digesta Flow. *J Dairy Sci* **69:** 413-419.

Sievert SM, Ziebis W, Kuever J, Sahm K (2000). Relative abundance of Archaea and Bacteria along a thermal gradient of a shallow-water hydrothermal vent quantified by rRNA slot-blot hybridization. *Microbiology* **146:** 1287-1293.

Singh UB, Verma DN, Varma A, Ranjhan SK (1977). The relationship between rumen bacterial growth, intake of dry matter, digestible organic matter and volatile fatty acid production in buffalo (Bos bubalis) calves. *Br J Nutr* **38:** 335-340.

Snelling TJ, Genç B, McKain N, Watson M, Waters SM, Creevey CJ *et al* (2014). Diversity and community composition of methanogenic archaea in the rumen of scottish upland sheep assessed by different methods. *PLoS ONE* **9**.

Snelling TJ, Wallace RJ (2017). The rumen microbial metaproteome as revealed by SDS-PAGE. *BMC Microbiol* **17:** 9.

Sniffen CJ, Robinson PH (1987). Protein and fiber digestion, passage, and utilization in lactating cows. Microbial growth and flow as influenced by dietary manipulations. *J Dairy Sci* **70:** 425-441.

Sollinger A, Schwab C, Weinmaier T, Loy A, Tveit AT, Schleper C *et al* (2016). Phylogenetic and genomic analysis of Methanomassiliicoccales in wetlands and animal intestinal tracts reveals clade-specific habitat preferences. *FEMS Microbiol Ecol* **92**.

Spor A, Koren O, Ley R (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* **9:** 279-290.

Stark L, Giersch T, Wünschiers R (2014). Efficiency of RNA extraction from selected bacteria in the context of biogas production and metatranscriptomics. *Anaerobe* **29:** 85-90.

Steinfeld H, Gerber P, Wassenaar TD, Castel V, de Haan C (2006). *Livestock's long shadow: environmental issues and options*. Food & Agriculture Org.

Stevenson DM, Weimer PJ (2007). Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl Microbiol Biotechnol* **75:** 165-174.

Stewart FJ, Ottesen EA, DeLong EF (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *Isme j* **4:** 896-907.

Stewart JB, Chinnery PF (2015). The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet* **16:** 530-542.

Sturm M, Schroeder C, Bauer P (2016). SeqPurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics* **17:** 208.

Sun YZ, Mao SY, Zhu WY (2010). Rumen chemical and bacterial changes during stepwise adaptation to a high-concentrate diet in goats. *Animal* **4:** 210-217.

Swain RA, Nolan JV, Klieve AV (1996). Natural variability and diurnal fluctuations within the bacteriophage population of the rumen. *Appl Environ Microbiol* **62:** 994-997.

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

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV *et al* (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:** 41.

Taxis TM, Wolff S, Gregg SJ, Minton NO, Zhang C, Dai J *et al* (2015). The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Nucleic Acids Res* **43:** 9600-9612.

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45:** D158-d169.

Thomas T, Gilbert J, Meyer F (2012). Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp* **2:** 3.

Thornton PK (2010). Livestock production: recent trends, future prospects. *Philos Trans R Soc Lond B Biol Sci* **365:** 2853-2867.

Toomey N, Monaghan A, Fanning S, Bolton D (2009). Transfer of antibiotic resistance marker genes between lactic acid bacteria in model rumen and plant environments. *Appl Environ Microbiol* **75:** 3146-3152.

Torok VA, Percy NJ, Moate PJ, Ophel-Keller K (2014). Influence of dietary docosahexaenoic acid supplementation on the overall rumen microbiota of dairy cows and linkages with production parameters. *Can J Microbiol* **60:** 267-275.

Tsai YC, Conlan S, Deming C, Segre JA, Kong HH, Korlach J *et al* (2016). Resolving the Complexity of Human Skin Metagenomes Using Single-Molecule Sequencing. *MBio* **7:** e01948-01915.

Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI *et al* (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat Genet* **48:** 1413-1417.

Tveit AT, Urich T, Svenning MM (2014). Metatranscriptomic analysis of arctic peat soil microbiota. *Appl Environ Microbiol* **80:** 5761-5772.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3:** e2527.

Uyeno Y, Kawashima K, Hasunuma T, Wakimoto W, Noda M, Nagashima S *et al* (2013). Effects of cellooligosaccharide or a combination of cellooligosaccharide and live Clostridium butyricum culture on performance and intestinal ecology in Holstein calves fed milk or milk replacer. *Livestock Science* **153:** 88-93.

VanRaden PM (2008). Efficient methods to compute genomic predictions. *J Dairy Sci* **91:** 4414-4423.

Vetrovsky T, Baldrian P (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* **8:** e57923.

Wallace RJ, Rooke JA, Duthie CA, Hyslop JJ, Ross DW, McKain N *et al* (2014). Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. *Sci Rep* **4:** 5892.

Wallace RJ, Rooke JA, McKain N, Duthie CA, Hyslop JJ, Ross DW *et al* (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* **16:** 839.

Wang L, Hatem A, Catalyurek UV, Morrison M, 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:** e78507.

Wang P, Qi M, Barboza P, Leigh MB, Ungerfeld E, Selinger LB *et al* (2011). Isolation of high-quality total RNA from rumen anaerobic bacteria and fungi, and subsequent detection of glycoside hydrolases. *Can J Microbiol* **57:** 590-598.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73:** 5261-5267.

Ward TL, Hosid S, Ioshikhes I, Altosaar I (2013). Human milk metagenome: a functional capacity analysis. *BMC Microbiol* **13:** 116.

Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A *et al* (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5:** 27.

Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B *et al* (2012). Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods* **91:** 80-88.

Williams YJ, Popovski S, Rea SM, Skillman LC, Toovey AF, Northwood KS *et al* (2009). A vaccine against rumen methanogens can alter the composition of archaeal populations. *Appl Environ Microbiol* **75:** 1860-1866.

Wimmer V, Albrecht T, Auinger HJ, Schon CC (2012). synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* **28:** 2086-2087.

Wood DE, Salzberg SL (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15:** R46.

Wright ADG, Auckland CH, Lynn DH (2007). Molecular diversity of methanogens in feedlot cattle from Ontario and Prince Edward Island, Canada. *Appl Environ Microbiol* **73:** 4206-4210.

Wu S, Zhu Z, Fu L, Niu B, Li W (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* **12:** 444.

Xia Y, Kong Y, Seviour R, Yang HE, Forster R, Vasanthan T *et al* (2015). In situ identification and quantification of starch-hydrolyzing bacteria attached to barley and corn grain in the rumen of cows fed barley-based diets. *FEMS Microbiol Ecol* **91:** fiv077.

Xiang R, McNally J, Rowe S, Jonker A, Pinares-Patino CS, Oddy VH *et al* (2016). Gene network analysis identifies rumen epithelial cell proliferation, differentiation and

metabolic pathways perturbed by diet and correlated with methane production. *Sci Rep* **6:** 39022.

Xue B, Wang LZ, Yan T (2014). Methane emission inventories for enteric fermentation and manure management of yak, buffalo and dairy and beef cattle in China from 1988 to 2009. *Agriculture, Ecosystems and Environment* **195:** 202-210.

Yeoman CJ, White BA (2014). Gastrointestinal tract microbiota and probiotics in production animals. *Annu Rev Anim Biosci* **2:** 469-486.

Yin YY, Liu YJ, Zhu WY, Mao SY (2014). Effects of Acarbose Addition on Ruminal Bacterial Microbiota, Lipopolysaccharide Levels and Fermentation Characteristics In vitro. *Asian-Australas J Anim Sci* **27:** 1726-1735.

Young JW (1977). Gluconeogenesis in cattle: significance and methodology. *J Dairy Sci* **60:** 1-15.

Yu PH, Lu LX, Fan H, Kazachkov M, Jiang ZJ, Jalkanen S *et al* (2006). Involvement of semicarbazide-sensitive amine oxidase-mediated deamination in lipopolysaccharide-induced pulmonary inflammation. *Am J Pathol* **168:** 718-726.

Yu Z, Morrison M (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36:** 808-812.

Yurkovetskiy L, Burrows M, Khan AA, Graham L, Volchkov P, Becker L *et al* (2013). Gender bias in autoimmunity is influenced by microbiota. *Immunity* **39:** 400-412.

Yutin N, Kapitonov VV, Koonin EV (2015). A new family of hybrid virophages from an animal gut metagenome. *Biol Direct* **10:** 19.

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18:** 821-829.

Zhang Z, Xu D, Wang L, Hao J, Wang J, Zhou X *et al* (2016). Convergent Evolution of Rumen Microbiomes in High-Altitude Mammals. *Curr Biol* **26:** 1873-1879.

Zhou M, Hernandez-Sanabria E, Le LG (2009). Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Appl Environ Microbiol* **75:** 6524-6533.

Zhou M, Hernandez-Sanabria E, Guan LL (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. *Appl Environ Microbiol* **76:** 3776-3786.

Zhou M, Chung YH, Beauchemin KA, Holtshausen L, Oba M, McAllister TA *et al* (2011a). Relationship between rumen methanogens and methane production in dairy cows fed diets supplemented with a feed enzyme additive. *J Appl Microbiol* **111:** 1148-1158.

Zhou M, McAllister TA, Guan LL (2011b). Molecular identification of rumen methanogens: Technologies, advances and prospects. *Animal Feed Science and Technology* **166-167:** 76-86.

Zhu W, Lomsadze A, Borodovsky M (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38:** e132.