Investigation of the associations between the gastrointestinal microbiome and animal host performance and health

Joana Patrícia do Jardim Gonçalves Lima

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

Royal (Dick) School of Veterinary Studies

2021

Declaration

I, Joana Lima, hereby declare that this thesis is of my own composition, that the work contained herein is my own except where explicitly stated otherwise in the text. The work reported in this thesis contains no material previously submitted for the award of any other degree or professional qualification.

Joana Lima

September 30th 2021

Abstract

The projected human population growth drives a worldwide increased pressure to improve animal production systems from both the economic and environmental perspectives. Bovine ruminants are one of the most interesting sources of high-quality protein, due to their ability to convert human-undigestible plant biomass and utilize it in the production of meat and milk. This ability is due to symbiotic associations with their rumen microbiome, i.e., the collection of microorganisms (bacteria, protozoa, fungi, and archaea) inhabiting the rumen, and their microbial genes (i.e., the metagenome). Whereas mammals do not produce the necessary enzymes to breakdown and digest complex polysaccharides in fibrous plants, bacteria, protozoa, and fungi are able to ferment these into volatile fatty acids, microbial proteins, and vitamins, which are utilized by the ruminant host for maintenance, growth, and development. An excess of hydrogen produced during the fermentation process is utilized by archaea, leading to the production of methane, a greenhouse gas with 28 times higher warming potential than carbon dioxide, which is released mostly by eructation into the environment, and thus contributing to climate change. This thesis focuses on the investigation of the host animal-microbiome symbiosis and how it impacts host animal performance traits, including appetite, growth, and feed conversion efficiency. We also analysed the symbiotic association with the host's health, by considering the impact of the presence of a parasitic nematode on the gastrointestinal microbiome of cattle.

In the first chapter of this thesis, I present an overview of the state-of-the-art knowledge, addressing why these microbiome-focused studies are crucial for the future development of more efficient bovines with lower environmental impact, summarizing the main findings in the field so far, and underlining the current challenges. The second chapter of this thesis includes a comparison between taxonomic compositions obtained from processing 16S rRNA amplicon sequences derived from caecum, colon, and faecal pig samples, using two different bioinformatics pipelines, the MetaGenome Rapid

Annotation using Subsystem Technology (MG-RAST) and the Quantitative Insights Into Microbial Ecology 2 (QIIME2). The results suggested that the microbiota profiles differ significantly according to the bioinformatics pipeline applied, with consequences on the subsequent statistical analyses; for example, at family level the richness and evenness of the samples was higher when samples were processed in QIIME2. We also compared different data cleaning and filtering methods (e.g., application of a minimum relative abundance threshold of 1%), which led to inconsistent results, depending on the pipeline used to identify microbial taxa. When using the whole datasets in partial least squares discriminant analyses (PLS-DA) to discriminate between sample collection sites, MG-RAST-derived data led to more accurate results, whereas QIIME2 was more accurate when a minimum relative abundance threshold was applied. This study was published in the Journal of Microbiological Methods in May of 2021.

This first study provides a substantial insight into the challenges of taxonomic characterization of samples using 16S rRNA sequencing data, which could be at least partly reduced by using whole metagenome sequencing data. The high resolution of whole metagenome sequencing methods provides the opportunity to identify functional microbial gene orthologs (e.g., using the Kyoto Encyclopedia of Genes and Genomes, KEGG), allowing for the comprehensive understanding of the biological and biochemical networks involved in the digestive processes closely associated with the host animal performance and health. Therefore, in the third chapter, we investigated rumen metagenome data derived from whole metagenome sequencing of rumen samples taken from beef cattle at slaughter, and their association with host performance traits including feed conversion ratio, average daily gain, residual feed intake and daily feed intake (FCR, ADG, RFI and DFI, respectively). Our analyses based on partial least squares (PLS) models identified sets of 20, 14, 17, and 18 microbial genes whose relative abundances explained 63, 65, 66, and 73% of the variation of FCR, ADG, RFI, and DFI, respectively. This research was the first to investigate the association between rumen microbial genes and beef cattle performance, and it provides very detailed information

about the functional role of the microbiome on cattle digestive processes. For example, microbial genes associated with cellulose and hemicellulose degradation, vitamin B12 synthesis, and amino acid metabolism were identified as biomarkers for enhanced feed conversion efficiency, whereas biomarkers for inefficient feed conversion were associated with functions such as pathogen lipopolysaccharide synthesis, cationic antimicrobial peptide resistance, and degradation of toxic compounds. This study was published in the journal Frontiers in Genetics in August of 2019.

Most studies focused on associations between the microbiome and the bovine host are based on taxonomic and metagenomic compositions derived from samples taken shortly after slaughter, but whether these samples are representative of the microbiome found in the rumen at different stages of the host animal's life is still unclear. In chapter 4, a study on the longitudinal stability of the rumen microbiome of beef cattle throughout the finishing growth phase is presented. Samples were collected from 20 animals before a nitrate- or oil-based additive was included in their basal diets (pre-additive), at the start, mid and end of a 56-day testing period (during which the performance traits FCR, ADG, DFI, and RFI were measured), after the animals left the respiration chamber in which they were tested individually for methane production, and at slaughter. Our results highlight that the microbiome compositions are stable throughout the finishing growth period, both at the microbial genera and microbial genes levels. We used partial least squares models to predict each performance measure (FCR, ADG, DFI, and RFI), and methane production (CH_4 production in g/day and CH_4 yield in g/kg dry matter intake) trait based on the microbiota and metagenomic datasets derived from each timepoint and compared the variable importance in projection (VIP) scores and the regression coefficients extracted from each PLS model. The results showed substantial consistency throughout the timepoints, indicating that data collected from any timepoint during the finishing growth phase would be suitable for prediction of host traits. Additionally, we found that rumen microbial biomarkers previously identified in other studies based on samples taken at slaughter are also informative to predict host performance traits based on

rumen samples taken at earlier stages of the animal's life, underlining not only the stability of the microbiome but also our confidence in the results from previous studies.

The gastrointestinal microbiome is not only highly associated with host performance traits, as shown in chapters 3 and 4, but also with host health traits. We investigated the influence of the presence of the abomasal parasitic nematode *Ostertagia ostertagi*, and of a vaccine against the nematode, on the rumen and caecum microbiomes of dairy cattle, and these studies are presented in chapters 5 and 6, respectively. A total of 24 calves were included in the experiment, of which 4 were left unvaccinated and uninfected throughout the whole experiment (UNF), 10 received a native vaccine against *O. ostertagi*, and 10 were injected with adjuvant only (positive control). The vaccinated and positive control groups were then subjected to the infection challenge that consisted of oral administration of 1000 infectious L3 larvae per day for 25 days. The animals were evaluated based on their cumulative faecal egg count (cFEC), leading to the identification of 4 vaccinated animals (with average cFEC, VAC) and 8 animals from the positive control group (4 extremely low and 4 extremely high cFEC, CLE and CHE, respectively) for whole metagenome sequencing and microbiome profiling at both taxonomic and microbial genes levels. The results indicate that the parasitism by this abomasal nematode substantially impacted the rumen and caecum microbiomes, and that the impact on the microbiome differs according to the parasitism severity. For example, in comparison to UNF, the rumen of CLE animals was depleted of microbial genes associated with valine, leucine, and isoleucine metabolism, whereas the rumen of CHE animals was depleted of microbial genes associated with bacterial chemotaxis and flagellar assembly. Additionally, in comparison to UNF, the rumen of VAC animals was enriched in microbial genes associated to metabolism of vitamin C and vitamin B12 synthesis. In the caecum of infected animals (CLE and CHE) we observed enrichment of several genera belonging to the phylum Actinomycetia, e.g., *Cellulomonas*, and depletion of several genera in phyla Gammaproteobacteria and Bacilli with pathogenic potential, such as *Legionella* and *Melissococcus*.

At the functional level, the analyses revealed enrichment of microbial genes associated with uptake of urea, and degradation of aromatic compounds in the caecum of infected, in comparison to uninfected animals. The rumen and caecum of vaccinated animals was enriched in several fungi, most of which with potential pathogenicity, e.g., *Colletotrichum* and *Botrytis* in the rumen, and *Pneumocystis* and *Malassezia* in the caecum, in comparison to uninfected animals.

The final chapter of the thesis includes a summary of the results of all previous chapters, and a general discussion on the symbiotic relationships between the gastrointestinal microbiome and host animal performance and health traits.

Lay Summary

The global population is expected to reach 9.8 billion by 2050, which will increase the pressure on food production systems to become more economically and environmentally sustainable. Livestock production is one of the most important food sources, with important implications in worldwide livelihoods, including food security and economic sustenance. The worldwide increased demand for high quality protein-rich food, such as meat and milk, emphasizes the need for added value, improved safety, minimised waste, and minimized environmental impact in livestock production systems.

Large ruminants, such as cattle, play a vital role in the global agricultural system, and since the 1960s, beef production has more than doubled. Ruminants can convert human-indigestible feed such as grass and silage, into quality foods, therefore contributing substantially to food security and economic development. Although mammals do not produce the enzymes necessary to digest the fibrous plant material ingested, ruminants have developed several strategies that aid in this process: on one hand, they are able to regurgitate the feed and re-chew it, increasing the mechanical breakdown process; on the other hand, the rumen hosts microbial organisms (bacteria, archaea, protozoa and fungi, i.e., the rumen microbiota) that digest the fibrous plant material through the enteric fermentation process (the fermentation process allows microbes to obtain energy for their own growth and activities from plant material in the absence of oxygen), releasing volatile fatty acids, microbial protein and vitamins, that the animal host uses for maintenance, development, and growth.

In this thesis, the gastrointestinal microbiome (which includes the microbiota, and their microbial genes, i.e., the metagenome) and its symbiotic relationship with the bovine host has been analysed with focus on the characterization of the microbiome profiles, and their influence on host animal performance (including appetite, growth, and feed conversion efficiency) and host health-related traits.

Although substantial research has been devoted to investigating the role of the microbiome in ruminal digestion, the rumen function is still under-characterized, mainly due to the complexity of the biochemical networks underlying the rumen functionality, built up on the high population density and high diversity of microorganisms. As a result of several investigative works, it has been established that the ruminant and its rumen microbiome exist and evolve in a symbiotic relationship, in which the host animal provides heat, moisture, and food for the microbes, whereas the rumen microbes ferment the ingested plant materials, rich in cellulose and hemicellulose (i.e., structural carbohydrates of the fibrous plant material), and release sub-products such as volatile fatty acids, source of energy and nutrients for the host animal and other microbes, as well as being themselves digested by the host animal as the main source of protein. The study of the microbiome in association with host animal traits is fundamental for the comprehensive understanding of the host animal characteristics, particularly those associated with performance, health, and even environmental impact, and it is expected to greatly contribute for the future development of improved animal production systems, for example by influencing diet composition formulation and selection of animals with desired characteristics in breeding programmes.

The complex microbial ecosystem composition within the rumen, and therefore the fermentation processes, are subject to the influence of many factors, including the age and development stage of the host, the size of the rumen, the feed passage rate, the host genetics, and diet composition. The host individual genetics explains a part of the variability observed in complex traits, such as feed conversion efficiency and health-related traits, which underlines even further the need for the characterization of the microbiome. Molecular techniques allow for a deeper understanding of the rumen microbiology, providing insight into the bacterial composition in the rumen without the need for culturing the microbes in the lab. Culture-independent methods carry a great advantage, given that around 80% of the rumen microbiota is thought not to grow using standard culture techniques. Furthermore, with the advent of next-generation sequencing techniques, it became possible to obtain the

microbial genetic composition of the microbiota (i.e., the metagenome), which aids in the interpretation of the functions of the microbiota in the rumen.

This thesis starts by providing a comprehensive summary of the research area, including the importance of ruminants as high-quality protein sources, and the importance of the host animal-microbiome interaction studies for the improvement of production systems. Subsequently, I summarized the most up-to-date literature to provide the reader with a comprehensive background on the rumen microbiome in association with the bovine host, and the major factors affecting the microbiome, such as diet composition and host genetics. Thereafter, I included a section on the influence of the rumen microbiome on host traits, including performance traits such as feed conversion efficiency, appetite, and growth, and health-associated traits such as parasitism by nematodes, one of the most economically impactful factors in grazing animals.

The second chapter includes a comparison between taxonomic compositions obtained based on a microbial gene that allows for the identification of different microbial groups, called the 16S rRNA gene. This gene exists in all microbes, and it varies sufficiently between different groups of microbes (i.e., genera), thus allowing for the characterization of each sample at the genera level. We extracted the 16S rRNA genes from caecum, colon and faecal pig samples, and processed them using two different bioinformatics pipelines, the MetaGenome Rapid Annotation using Subsystem Technology (MG-RAST) and the Quantitative Insights Into Microbial Ecology 2 (QIIME2), thus obtaining the taxonomic composition of each sample. These tools process the samples in different manners; whereas MG-RAST calculates the similarity between all the 16S rRNA sequences to group them into Operational Taxonomic Units (OTUs), QIIME2 first identifies unique sequences (i.e., 16S rRNA gene amplicons that have exactly the same sequence), and then uses the abundance of each unique sequence and its similarity to other unique sequences to group them into Amplicon Sequence Variants (ASVs). We found significant differences between the microbiota profiles obtained from each pipeline; for example, we observed higher evenness and richness at family

level when samples were processes in QIIME2. Microbiota-focused studies rely on the accurate taxonomic characterization of the samples, and nowadays, researchers are very focused in finding how the taxonomic composition of samples relates to the host animal traits. Thus, with the goal of verifying the implications of using different pipelines to obtain the taxonomic composition, we investigated how the taxonomic compositions obtained from each pipeline were associated to the sampling collection site, i.e., we analysed what were the main differences between using the taxonomic compositions obtained from MG-RAST and QIIME2 to distinguish samples collected from the caecum, colon, and faeces of pigs. Our results suggested that using the taxonomic compositions of the samples as processed by different bioinformatics pipelines would lead to different statistical results, and for example, when using MG-RAST, the microbes considered more important for distinguishing the sample collection sites were *Acetitomaculum*, *Ruminococcus* and *Methanosphaera*, whereas when using QIIME2, these were *Candidatus Methanomethylophilus*, *Sphaerochaeta* and *Anaerorhabdus*.

In the third chapter, we explored the metagenome of beef cattle from four economically significant breeds (Aberdeen Angus, Charolais, Luing and Limousin), fed two basal diets differing in fibre content, in association with the host animal performance traits, including feed conversion ratio (FCR), average daily weight gain (ADG), residual feed intake (RFI) and daily feed intake (DFI). This research indicated that there is a substantial link between the rumen microbial genes and the host animal performance, identifying sets of 20, 14, 17, and 18 microbial genes whose relative abundances explained 63, 65, 66, and 73% of the variation of FCR, ADG, RFI, and DFI, respectively, and many functional explanations for these associations were elucidated. For example, we identified microbial genes with functions such as cellulose and hemicellulose degradation, vitamin B12 synthesis and amino acids metabolism as biomarkers for high feed conversion efficiency, whereas feed-conversion-inefficient animals were characterized by high relative abundance of microbial genes associated with nucleotide sugars metabolism, pathogen

lipopolysaccharide synthesis, cationic antimicrobial peptide resistance and degradation of toxic compounds.

Chapter four is dedicated to the analyses of the rumen microbiome of beef cattle throughout the finishing growth phase (adult animals). Rumen microbial samples were taken from 20 animals before receiving diet additives (nitrate or oil based), at the start, middle and end of a 56-day testing period (during which the performance traits FCR, ADG, RFI and DFI were measured), after leaving the respiration chamber in which the animals were individually measured for methane emissions) and at slaughter. Our results indicate that the microbiome compositions remained highly stable throughout the finishing period. Additionally, the microbiome compositions at microbial taxa (i.e., genera) and microbial genes levels at each timepoint were used to explain the variation observed in FCR, ADG, DFI, RFI, methane production and methane yield, and showed that the association between microbes and microbial genes with each performance and methane emissions trait was stable through time.

The fifth and sixth chapters present the results from an investigation of the influence of the presence of the abomasal parasitic nematode *Ostertagia ostertagi* on the rumen and caecum microbiomes, respectively, in dairy cattle. The influence of a vaccine against this nematode on the microbiomes was also explored. This study included 16 calves, 4 were left unvaccinated and uninfected throughout the whole experiment, and 12 were subject to an infection challenge consisting of oral administration of 1000 L3 larvae/day for 25 days, of which 4 had previously been vaccinated against the nematode. The rumen and caecum microbiomes were substantially influenced by the presence of the nematode, for example, the rumen of infected animals was depleted of microbial genes associated to folate, amino acids and fatty acids biosynthesis and metabolism. Additionally, the vaccine was also shown to influence rumen microbiome compositions, with vaccinated animals showing enrichment of microbial genes associated to vitamin C metabolism and vitamin B12 biosynthesis. Regarding the caecum, we reported enrichments of actinomycetes and depletion of Gammaproteobacteria and Bacilli in infected

in comparison to the uninfected animals. Several fungi (Ascomycota and Basidiomycota) were enriched in the rumen and caecum microbiome profiles of vaccinated animals, most of them were opportunistic pathogens.

The last chapter summarizes all results obtained throughout the PhD project and concatenates them into a general discussion on the microbiome in association with different host traits, including those related to performance and health.

Acknowledgements

This thesis is a result of a collaborative project between Scotland's Rural College (SRUC), and the University of Edinburgh. I am grateful to SRUC, European Union, BBSRC, The Scottish Government, Zoetis, AHDB, Defra, and QMS, who directly or indirectly funded my PhD project. I would also like to extend my gratitude to all the people involved in the collection of the data with which I worked.

I would like to thank my supervisors, Rainer Roehe, Richard Dewhurst, and Tom Freeman, for giving me this great opportunity. I am extremely thankful that you took this chance on me, and very grateful for all your support during these last 4 years. I have learned a lot, and your guidance was crucial for my development, not only as a scientist, but also as a person.

I would also like to thank my team, Rainer Roehe, Tuan Nguyen, Marina Martinez-Alvaro, and Marc Auffret, each with different interests, and areas of expertise have helped me throughout this project. Rainer was an excellent leader, with great dedication and passion for the science, and always received me with patience and open-mindedness. Marina and Marc were excellent, albeit very different, examples of work discipline, and scientific curiosity. But most of all, I want to thank Tuan, who was my companion from day one, and without whom I would most certainly not have had the courage to see this whole project through. I also want to thank you all for the fun moments we had, whether it was picnics in the Meadows, nights out in Aberdeen, or beers in Ghent, I always felt I was among friends.

During this PhD, I also made many friends, which made this journey a much more fun and exciting one! I would like to thank Harry Kamilaris, Fran Shepherd, Alex Morris, Laura Salazar, Marie Rowland, Jay Burns, Cathrine Erichsen, Lucy Oldham, Miguel Somarriba, Riccardo Bica, Antonio Pacheco, Mengyuan Wang, and our dearly missed Aluna Chawala, remembering all our lunch and coffee breaks, little walks around the building, and obviously our amazing Writing Retreat! I could not have asked for better company!

A big thank you to my family in Portugal, most of all to my mum Luciana Carvalho, who never gave up on trying to make me a better and harder worker, and to my dad, António Lima. I also want to thank my brother and best friend, João Lima, who always believed in me, even when I didn’t. My thank you extends also to both my godmothers, Maria Júlia Reis, and Helena Dias, not only for the help during my MSc and my jump from Portugal to Scotland, but also for their close involvement in my education and growth throughout all my life.

Finally, for his love, support, and especially the caring way in which he dealt with the rollercoaster of emotions throughout the final stretch of this project, I would like to deeply thank my partner Antonis Mastoras.

Abbreviations

ADG	Average Daily Weight Gain
ASV	Amplicon Sequence Variant
CH_4	Methane
CO_2	Carbon Dioxide
DFI	Daily Feed Intake
DMI	Dry Matter Intake
DNA	Deoxyribonucleic Acid
DGGE	Denaturating Gradient Gel Electrophoresis
FCR	Feed Conversion Ratio
GIT	Gastrointestinal Tract
GHG	Greenhouse Gas
H_2	Hydrogen
KEGG	Kyoto Enciclopedia of Genes and Genomes
MG-RAST	Metagenome Rapid Annotation using Subsystem Technology
NDF	Neutral Detergent Fibre
n-RFI	Negative Residual Feed Intake
OTU	Operational Taxonomic Unit
p-RFI	Positive Residual Feed Intake
PCR	Polymerase Chain Reaction
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis

QIIME	Quantitative Insights Into Microbial Ecology
QIIME2	Quantitative Insights Into Microbial Ecology 2
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
VIP	Variable Importance in Projection
VFA	Volatile Fatty Acids
WGS	Whole Genomic Sequencing

List of Publications

Lima, J., Auffret, M.D., Stewart, R.D., Dewhurst, R.J., Duthie, C-A, Snelling, T.J., Walker, A.W. Freeman, T.C., Watson, M., Roehe, R. (2019) Identification of Rumen Microbial Genes Involved in Pathways Linked to Appetite, Growth, and Feed Conversion Efficiency in Cattle. Frontiers in Genetics. 10:701. 10.3389/fgene.2019.00701

Lima, J., Manning, T., Rutherford, K. M., Baima, E. T., Dewhurst, R. J., Walsh, P., & Roehe, R. (2021). Taxonomic annotation of 16S rRNA sequences of pig intestinal samples using MG-RAST and QIIME2 generated different microbiota compositions. Journal of Microbiological Methods, 186(March), 106235. https://doi.org/10.1016/j.mimet.2021.106235

Conference Proceedings

Lima, J., Auffret, M.D., Stewart, R.D., Dewhurst, R.J., Duthie, C-A, Walker, A.W., Freeman, T.C., Watson, M., Snelling, T.J., Roehe, R. (2019). Investigating the impact of rumen microbial genes on feed conversion efficiency, growth rate and feed intake in beef cattle. The Proceedings of the British Society of Animal Science - BSAS 75th Annual Conference 2019. Edinburgh, Scotland, 9-11 April 2019.

Lima, J., Auffret, M.D., Stewart, R.D., Dewhurst, R.J., Duthie, C-A, Snelling, T.J., Walker, A.W., Freeman, T.C., Watson, M., Roehe, R. (2019). Metagenomic analysis indicates an association of rumen microbial genes with appetite in beef cattle. Book of Abstracts of the 70th Annual Meeting of the European Federation of Animal Science (EAAP). Ghent, Belgium, 26-30 August 2019

Lima, Joana, McNeilly, T., Steele, P., Martinez-Alvaro, M., Auffret, M., Dewhurst, R., Roehe, R., & Watson, M. (2021). Altered rumen microbiota profiles in cattle infected with and vaccinated against the parasitic nematode Ostertagia ostertagi. Animal - Science Proceedings, 12(1), 108. https://doi.org/10.1016/j.anscip.2021.03.130

Contents

Chapter 1 General introduction

1.1 Introduction

The United Nations estimated the world population in 2019 at 7.7 billion people, with an expected rise to 9.7 billion people by 2050 (Department of Economic and Social Affairs Population Division - United Nations, 2019). The populational growth observed during the past few centuries and estimated for the near future increases the need for ever more efficient food production systems, in both economic and ecological perspectives, mainly due to an escalation in the global demand for food. Livestock production systems are subject to different pressures, depending on their socioeconomic context. Whereas in developing countries the main goal is to increase livestock products' availability for the consumer, in developed countries the focus is on increasing product quality and production efficiency through the implementation of more environmentally sustainable, and animal welfare-aware strategies (Thornton, 2010).

Meat is the most valuable livestock product, mainly due to its nutritional value – it provides high quality protein, all essential amino acids, and it is rich in minerals and vitamins. Whereas in developed countries meat consumption has stabilized (at high levels), the global meat market is expected to double its output by 2050, mostly driven by an escalating demand in developing countries, due to growing population and increasing incomes (Thornton, 2010; Food and Agriculture Organization (FAO), 2019). Livestock products are therefore an extremely important commodity, as reflected in the more than doubled production of beef and the more than 10-fold increase in chicken meat production, coupled with increased carcass weight and milk and eggs production per animal by about 30% since the 1960s (Thornton, 2010).

The pressure for increased livestock production has some obvious drawbacks associated with overexploitation and pollution; for example, the extensive use of land and water to maintain and feed animals is directly associated with loss of biodiversity and natural habitats, soil degradation, and water pollution.

Additionally, worldwide food consumption patterns from 2007 to 2016 were responsible for 21 to 37% of the total anthropogenic greenhouse gas (GHG) emissions, significantly contributing to climate change (FAO, 2020). The enormous challenge involved in achieving worldwide food security will require extreme efforts, not only through the increased efficiency of livestock and agricultural production systems while mitigating GHG emissions, water, and land use, but also through socioeconomic and political considerations and informed actions (Godfray et al., 2010).

Providing more than half of all protein in the livestock sector, ruminants occupy a prominent position in food security. Ruminants are extremely interesting because of their ability to feed on forages, agro-industrial by-products, and crop residues that would otherwise be wasted. Not only they provide us with high quality protein (i.e., milk and meat), they are also a source of manure used for soil fertilization, and they are essential for the livelihoods of millions of farmers worldwide (Food and Agriculture Organization (FAO), 2021).

In this chapter, an overview of the ruminant – particularly bovine cattle - is presented, together with different strategies for increased meat and milk productivity. The ruminant genetic make-up, its rumen microbiome, and environmental factors are focused on due to their widely recognized impact on the animals' performance traits, such as appetite, growth rate, feed conversion efficiency, and methane (CH_4) emissions, and on health traits, such as resilience to parasites.

1.2 Ruminants

Ruminant animals are herbivore ungulate mammals characterized by their rechewing of the cud. This group includes bovine cattle, sheep, goats, bison, antelopes, giraffes, gazelles, and camelids. Ruminants can utilize a vast range of plant-based resources, by fermenting the feed in a specialized compartment – the rumen. The rumen is the first of four chambers in the digestive tract of ruminants. It has a volume capacity in an adult bovine above 100L, with pH ranging from 5.7 and 7.3, and temperature around 39^{o} C. The rumen wall is a

major site for nutrient absorption, and it is covered in papillae which increase the inside surface of the rumen. These characteristics make the rumen an ideal fermentation chamber (Moran, 2005). The rumen is inhabited by an array of bacteria, archaea, protozoa, and fungi, which maintain complex interrelationships (e.g., commensalism, mutualism, competition, and predation). The symbiotic relationship between the ruminant animal and the microbial communities inhabiting its rumen is the main reason cattle can feed on plant-based, even if low-quality, feedstuffs, by utilizing the fermentation products for their growth and development without directly competing with humans for resources.

1.3 Microbiome and microbiota – important definitions

Berg et al. (2020) reviewed multiple definitions of the term *microbiome*, and finally proposed the following:

> *The microbiome is defined as a characteristic microbial community occupying a reasonable well-defined habitat which has distinct physio-chemical properties. The microbiome not only refers to the microorganisms involved but also encompass their theatre of activity, which results in the formation of specific ecological niches. The microbiome, which forms a dynamic and interactive micro-ecosystem prone to change in time and scale, is integrated in macro-ecosystems including eukaryotic hosts, and here crucial for their functioning and health.*

Additionally, the authors proposed the definition of *microbiota* as follows:

> *The microbiota consists of the assembly of microorganisms belonging to different kingdoms (Prokaryotes [Bacteria, Archaea], Eukaryotes [e.g., Protozoa, Fungi, and Algae]), while "their theatre of activity" includes microbial structures, metabolites, mobile genetic elements (e.g., transposons,*

phages, and viruses), and relic DNA embedded in the environmental conditions of the habitat.

Following the definitions presented above, throughout this thesis the rumen microbiome refers to the microbial communities (i.e., bacteria, archaea, protozoa, and fungi) and all their microbial genes, as well as metabolites, and physio-chemical properties of the rumen as a well-defined environment, whereas the rumen microbiota refers solely to the assembly of microbial organisms found in the rumen environment.

1.4 The rumen microbiome in the bovine host

The rumen microbiome is closely associated with the animal host's performance traits (e.g., appetite, growth rate, and feed conversion efficiency), mostly because the utilization of plant-based feed by the animal host is only possible due to the fermentation by the rumen microbiota of compounds that are undigestible by mammalian enzymes.

The rumen microbiota includes microbial communities with cellulolytic, hemicellulolytic, and pectinolytic capabilities, which anaerobically ferment complex polysaccharides cellulose, hemicellulose, and pectin, respectively, into monosaccharides. These monosaccharides are then converted into volatile fatty acids (VFAs), carbon dioxide (CO_2) and hydrogen (H_2), mostly by acetogenic and acidogenic bacteria (Ahring et al., 2018).

Although the rumen microbiota profiles differ between ruminant species, diets, and geographical location (among other factors), there is a core microbiota (i.e., a group of microorganisms that are virtually present in all ruminants) of dominant bacteria, which includes microbial genera *Prevotella*, *Butyrivibrio*, and *Ruminococcus*, microbial families *Lachnospiraceae* and *Oscillospiraceae* (i.e., *Ruminococcaceae*), and microbial orders *Eubacteriales* and *Bacteroidales* (Henderson et al., 2015).

Of all the microbial communities identified in the rumen, *Prevotella* is certainly the most commonly mentioned; in a 16S rRNA gene-based meta-analyses of

2662 bacterial samples collected from the rumen of dairy and beef cattle from a total of 53 previous studies performed in 29 countries, Holman & Gzyl (2019) identified *Prevotella* at a prevalence of over 95% and relative abundance averaging 24.1%. *Prevotella* is an anaerobic Gram-negative, non-spore forming, non-motile, rod-shaped genus belonging to the phylum *Bacteroidetes,* that feeds mostly on starch, fructan, and/or hemicelluloses. This genus encompasses many different species (7 of which have been described in the rumen and hindgut), most of which are able to utilize plant storage carbohydrates fructan and starch, whereas others can grow on hemicelluloses such as xylan. *P. ruminicola* (one of the most abundant *Prevotella* species in the rumen) has been shown to be a generalist, able to grow on starch, inulin (fructan), beta glucan, xyloglucan and arabinoxylan; in contrast, *P. brevis* strains are specialists that rely only on starch and fructan (Accetto and Avguštin, 2019). The sole observation of the species within the *Prevotella* genus hints into the complexity of the associations between microbial communities in the rumen ecosystem. On the one hand, the proportion of plant polysaccharides in the feed suggests a certain degree of functional redundancy (i.e., many species being able to degrade the same substrate), reflecting the competitiveness of the environment; on the other hand, the diversity of these nutrients alludes to the possibility for specialization of the enzymatic systems found in microbial genomes, reflecting a coordinated ecosystem supported on a somewhat functional complementarity.

The genus *Butyrivibrio* belongs to the family *Lachnospiraceae*, class Clostridia, phylum Firmicutes, and includes anaerobic cellulolytic Gram-negative curved rods, although *B*. *fibrisolvens* has in their cell wall some Gram-positive-like structures including glycerol teichoic acids and lipoteichoic acids (Cheng and Costerton, 1977). In a comparative genetic analysis, Palevich et al. (2020) investigated the genomes of 30 *Butyrivibrio* strains, and identified a core genome composed of 2% of the collective genome, mainly associated with survival in the rumen (and housekeeping of the cells), suggesting that a large proportion of the collective genome is composed of unique strain-specific

genes, which confer the ability to occupy several different niches within the rumen ecosystem. *Butyrivibrio* species are metabolically versatile, with most strains showing the ability to grow on pectin and xylan, but not on cellulose.

As with *Butyrivibrio*, *Ruminococcus* is also a *Firmicutes*, in the Clostridia class, but it belongs to the *Oscillospiraceae* (i.e., *Ruminococcaceae*) family. The genus *Ruminococcus* is composed of Gram-positive anaerobes that ferment cellulose into acetate, formate, and succinate (Kaars Sijpesteijn, 1951; van Gylswyk and Labuschagne, 1971; Gokarn et al., 1997; Russell et al., 2009).

In a global census of ruminal communities of 32 ruminant and camelid species, archaea in the core microbiota were shown to be remarkably similar between species from different areas of the world. The core microbiota included methanogenic archaea such as *Methanobrevibacter gottschalkii* and *Methanobrevibacter ruminantium*, and methylotrophs such as *Methanomassiliicoccales* (Henderson et al., 2015). These organisms have hydrogenotrophic and methylotrophic ability, whereas they can utilize hydrogen and methyl groups derived from methanol and methylamines, respectively, in their growth, with consequent production of methane.

Rumen protozoa can account for up to 50% of the biomass in the rumen. These organisms have fibrolytic capabilities and are involved in carbohydrates transport and metabolism (Williams et al., 2020). Protozoa *Entodinium* and *Epidinium* are highly abundant in the rumen (Henderson et al., 2015).

Rumen fungi (also known as gut fungi) such as *Neocallimastix*, *Piromyces* and *Orpirzomyces*, are anaerobic organisms that produce hydrolases, mostly organized in cellulosomes (uniquely found in gut fungi). These enzymatic complexes are extremely active in the degradation of cellulose, but gut fungi are also able to utilize xylan, mannose, ester, starch, and glucan (Fliegerova et al., 2015). During fermentation, rumen fungi produce formate, acetate, lactate, ethanol, CO_2, and H_2 (Akin and Borneman, 1990). Rumen fungi communities were found to be enriched in the rumen of dairy cattle fed a high-fibre diet, and this was associated with their fibrolytic activity, as well as with

enrichment of *Ruminococcaceae* and *Succiniclasticum*, suggesting an interaction between bacterial groups and rumen fungi (Kumar et al., 2015).

1.5 Diet composition influences the rumen microbiome profiles

The major functions of rumen microbes and their interactions which are addressed in this thesis have been mostly identified by dietary interventions and are therefore reviewed in this section. By providing different substrates, diets differing in nutritional content will promote and/or inhibit the growth of specific groups of microorganisms, each with their own growth requirements and fermentation products. Additionally, the diet composition influences the microbial enzymatic activities. For example, the microbiome structure has been shown to shift through increased growth of amylolytic and other starch-digesting microorganisms when animals are adapted from forage-based to concentrate-based diets (Fernando et al., 2010). Additionally, the complementation of forage-based feed with concentrate has been shown to decrease the activity of fibrolytic enzymes of the microbiota adhering to the plant particles (Nozière et al., 1996).

The primary component of cattle diet is forage, i.e., roughage, like silages, hay, straw, and grass; this is often complemented by some type of concentrate (usually based on grains, such as barley, or co-products from the food and drink industries) in varying proportions, depending on the production system, as well as fats.

Forage is rich in fibre, i.e., it contains a large proportion of plant cell wall water-insoluble carbohydrates such as cellulose, hemicellulose, and lignin, and cell wall water-soluble carbohydrates, including pectin. Mammals do not produce the enzymes necessary to digest fibre, but some microbial organisms harboured in their digestive tract do, and that is why the rumen microbiota (in the case of foregut fermenters such as cattle) is paramount for the digestive function in ruminants. The fibrous content in feed is mainly digested by cellulolytic and fibrolytic microorganisms in the rumen. Ruminal digestion of

water-insoluble carbohydrates takes up to 10 hours and this lag phase has been associated with the time it takes for feed particles to be broken down into smaller particles and colonized by microbial organisms (Nozière et al., 2010). The water-insoluble fibre that escapes ruminal digestion will pass into the small intestine and reach the caecum, where it will be fermented by the resident microbial communities (Michalet-Doreau et al., 2002; Nozière et al., 2010).

During ruminal fermentation, cellulose and hemicellulose are mainly broken down into glucose, glucose-6-phosphate, fructose-6-phosphate, and triosephosphate by enzymes including cellulase. Pectins and part of the hemicellulose are broken down into xylose and other pentoses. These are then degraded into hexoses, which in turn are converted into fructose-6-phosphate and triosephosphates.

The concentrate component of the diet usually has starch as its main component. Since the saliva of ruminants does not contain amylase, this will mainly be degraded in the rumen by microorganisms with amylolytic capability (Nozière et al., 2010). Compared to water-insoluble carbohydrates, starch is rapidly digested by the rumen microbiota (Nozière et al., 2010). Overall, increased availability of rapidly fermentable carbohydrates contributes to increased ruminal microbial fermentation and consequent increased production of VFA, mostly by bacteria and protozoa groups (Nagaraja and Titgemeyer, 2007). Storage carbohydrates such as starch, dextran, and simple carbohydrates (mono- and disaccharides) are broken down by microorganisms into maltose, and then converted into glucose-1-phosphate.

The products from the microbial fermentation of water-soluble and -insoluble carbohydrates are then used as substrates in glycolysis, producing pyruvate, which is then converted into acetate, propionate, and butyrate, the main VFAs produced in the rumen. VFAs are the most important energy source for the ruminant animal, being readily absorbed through the rumen wall, and serving as energy for the animal’s tissues.

Along with VFA, ammonia is one of the most important products of fermentation (Moran, 2005). Ammonia is utilized as nitrogen source for production of microbial protein, during the growth and reproduction of the rumen microbiota. Microbial protein is the most important protein source for the animal, as it supplies the animal with up to 80% of its amino acid requirements (Loor et al., 2016).

Since diet composition greatly influences the rumen microbiota, diet composition can be used as a tool to manipulate the microbiota.

The increased content of rapidly fermentable substrates such as starch in high-grain diets compared to high-fibre diets promotes the growth of amylolytic bacterial groups, including *Bacteroides* and ciliate protozoa, such as entodiniomorphs, which engulf and slowly digest starch (Mackie et al., 1978; Goad et al., 1998; Tajima et al., 2001). Additionally, amylodextrin-, and maltose-utilizing bacteria including *Bifidobacterium*, *Butyrivibrio*, *Eubacterium*, *Lactobacillus*, *Mitsuokella*, *Prevotella*, *Ruminobacter*, *Selenomonas*, *Streptococcus*, *Succinimonas*, and *Succinivibrio*, are also promoted due to amylodextrins and maltose present in the rumen (Nagaraja and Titgemeyer, 2007), although *Eubacterium ruminantium* and *Succinivibrio dextrisolvens* have been reported to decrease in animals fed high-grain diets (Tajima et al., 2001). In contrast, the growth of fybrolytic organisms, such as *Ruminococcus flavefaciens* and *Fibrobacter succinogenes*, and xylanolytic organisms is inhibited, reduced, or slowed down in animals fed concentrate diets, mostly due to their sensitivity to even mildly acidic environments (Tajima et al., 2001; Krause et al., 2003; Nagaraja and Titgemeyer, 2007). *Butyrivibrio fibrisolvens* is mainly a fibrolytic bacterium but it is also able to use maltose and sucrose. Although it can use both fibre and starch, it is usually decreased in the microbiomes collected from the rumen of animals fed high-grain diets, most likely due to the lower pH (Fernando et al., 2010).

The fermentation of starch and soluble sugars by the rumen microbiome quickly increases the concentration of VFAs and lactate in the rumen. The

increased lactate concentration in the rumen favours the growth of lactate-utilizing bacterial organisms, such as *Lactobacillus*, *Veillonellaceae* and *Selenomonas* (specifically *S. ruminantium ssp. lactilytica*), *Fusobacterium necrophorum*, *Megasphaera elsdenii*, *Peptostreptococcus asaccharolyticus*, *Propionibacterium acnes* (Mackie et al., 1978; Goad et al., 1998; Tajima et al., 2001; Nagaraja and Titgemeyer, 2007), which metabolize lactate into butyrate and propionate. The accumulation of VFA leads to increased acidity in the rumen, which will inhibit the growth of these groups (including ciliate protozoa) in favour of bacterial organisms with acid-tolerance, such as *Anaerovibrio* spp. and *Streptococcus bovis* (Mackie et al., 1978; Russell and Hino, 1985; Nagaraja and Titgemeyer, 2007).

The marked decrease in pH observed in the rumen of animals fed high-grain diets compared to high-fibre diets has been associated with the increased VFA production, rather than the lactic acid accumulation, most likely due to increased growth of lactate-utilizing bacteria (Goad et al., 1998). For example, *Megasphaera elsdenii* is a lactate-utilizing butyrate-producing bacterium that is more abundant in animals fed grain-rich diets in comparison to hay-fed animals, whereas butyrate production may serve as an electron sink during oxidation of lactate to pyruvate (Fernando et al., 2010). Additionally, although *Streptococcus bovis* is a mixed acid fermenter able to switch to homolactic fermentation if the pH is below 5.6, increasing the production of lactate and consequently the acidity in the rumen, its growth rate at pH lower than 6 decreases considerably, whereas the growth rate of *Lactobacilli* increases, and an antagonistic relationship between these taxa has been proposed (Nagaraja and Titgemeyer, 2007). The low pH in the rumen can lead to digestive disorders, such as acidosis, with a negative impact on the animals' performance (Bergman, 1990). Therefore, the adaptation of the microbiota communities towards greater resistance to lower pH may be regarded as a defence mechanism against acute acidosis, underlining the symbiotic association between the animal host and its rumen microbiome. Adaptation of animals to high-grain diets from high-fibre diets is usually performed in a

progressive manner, with gradual increases in the grain content of the diet and decreases in the fibre content.

Additionally, rumen microbiome profiles exhibit lower alpha-diversity when animals are fed high-grain diets in contrast with high-fibre diets (Anderson et al., 2016). These examples underline the complexity of the rumen microbiome, and the multitude of factors influencing associations between microbes themselves and the ruminal environment. Although these studies, among others, have been pivotal for the current understanding of the fluctuations of microbial groups within the microbiome due to changes in diet composition, most of them involved a low number of animals, and/or were based on culture-dependent methods, providing limited insights into the rumen microbiome processes and dynamics.

1.6 Animal host genetics influence rumen microbiome profiles

Host genetics have been shown to strongly influence the microbiome composition. Roehe et al. (2016) used sire progeny groups to estimate the genetic influence of bovine hosts on their methane emissions. The results suggested that the relative abundances of archaea in the rumen were under genetic influence of the host, and highly associated with methane emissions, suggesting the microbiome to be, at least partly, controlled by the host genetics. Difford et al. (2018) reported that bovine host genetics determined part of the variation in the relative abundance of some bacterial and archaeal organisms in the rumen, and that inter-individual differences in CH_4 production were associated with different genetic background and rumen microbiome composition. In a study that included 709 beef cattle, Li et al. (2019) estimated the heritability of taxonomic bacterial and archaeal groups (taxonomy was resolved based on 16S rRNA gene), and found moderate heritability estimates for taxonomic groups associated with feed conversion efficiency and rumen VFAs. More recently, Martínez-Álvaro et al. (2021) applied whole metagenome sequencing to derive microbial genetic and taxonomic profiles from the rumen of 363 steers. From a total of 1107 microbial genera and 1141 microbial genes

identified, 203 and 352, respectively, had significant moderate to high heritability, confirming the influence of the host genetics on their own microbiome profiles. The host genetic effect on microbiome features is of great interest, particularly due to its potential application in selecting animals for breeding that carry the rumen microbiome profiles associated with, for example, higher feed conversion efficiency or with the highest resilience to infection by nematodes.

1.7 Host performance traits

Feed conversion efficiency is a biologically complex trait, and it is influenced by several other traits, including feed intake, absorption rate, activity of the animal, and maintenance requirements. Feed conversion efficiency is often assessed through feed conversion ratio (FCR) and/or residual feed intake (RFI). Whereas FCR is calculated as the ratio between the average daily feed intake (DFI) and the average daily weight gain (ADG) of an animal, RFI corresponds to the difference between the observed DFI and the expected DFI, which is estimated by a multiple regression of observed DFI on potential energy sinks, typically including maintenance, lean and fat tissue growth, and activity. Due to the high cost of feeding production animals, the improvement of feed conversion efficiency is extremely attractive from an economic perspective.

FCR is calculated as a ratio, which implies that changes in FCR may be due to either changes on the denominator (DFI) and/or on the numerator (ADG). This means that genetic selection for lower FCR may lead to differential selective pressure on the component traits. In contrast, RFI is, at phenotypic level, independent of the traits that are fitted in the regression model used to estimate the expected DFI (Berry and Crowley, 2013). However, independence at phenotypic level does not warrant genetic independence. In fact, Arthur et al. (2001) estimated the genetic correlation between FCR and RFI at 0.66 ± 0.05, showing that these are at least partly different traits. Additionally, Arthur et al. (2001) estimated a lower heritability for FCR than for RFI, corresponding to 0.29 and 0.39, respectively.

Feed conversion efficiency traits are complex, not only because they are difficult to measure, but also due to their biological association with so many other traits. In a meta-analysis of 9 published studies on growing beef cattle, Kenny et al., (2018) reported that high-RFI animals spend a significantly higher amount of time (10.3 minutes on average) eating per day than their low-RFI counterparts, showing that the appetite and feeding behaviour are closely associated with feed conversion efficiency. Additionally, variation in feed conversion efficiency may be influenced by factors such as intestinal absorption, as suggested by significant associations between jejunal, duodenal and ileal epithelial tissue density and RFI in cattle (Montanholi et al., 2013; Meyer et al., 2014). Higher feed efficiency (lower RFI) has been associated with reduced hepatic lipid synthesis and fat accumulation, indicative of a more efficient use of consumed energy and nutrients, potentially towards protein and lean muscle synthesis (Mukiibi et al., 2018).

1.8 The gastrointestinal microbiome influences the feed conversion efficiency of the bovine host

The rumen microbiome has been shown to be associated with feed conversion efficiency variation in animals. Carberry et al. (2012) analysed the association of specific rumen microorganisms with feed conversion efficiency and diet in beef cattle, using PCR-denaturating gradient get electrophoresis (DGGE) and quantitative PCR (qPCR). The authors suggested that the association between the rumen microbiota and RFI is dependent on the diet, in agreement with the RFI-based re-ranking of animals when these were adapted from a grower to a finisher diet (Durunna et al., 2011). McCann et al. (2014) investigated the association of the rumen microbiota and RFI of Brahman bulls on bermudagrass pastures, based on 16S rRNA amplicon data. The authors determined the core microbiota within positive- and negative-RFI animals (p-RFI and n-RFI, respectively); four *Ruminococcaceae*, five *Prevotella* and one *Paludibacter* operational taxonomic units (OTUs) were exclusively observed in the core microbiota of the n-RFI group, whereas ten *Prevotella*, three *Lachnospiraceae*, two *Oscillospira*, and one *Treponema* OTUs were

exclusively observed in the core microbiota of the p-RFI group. Furthermore, *Prevotellaceae* family was significantly relatively more abundant in the p-RFI group (McCann et al., 2014). In 2016, Li et al. (2016) further clarified the association of the rumen microbiota profiles with the host animal RFI. Regarding the microbiota, bacterial families *Lachnospiraceae*, *Veillonelaceae*, and p-2534-18B5 were found to be relatively more abundant in low efficiency (i.e., higher RFI) animals, whereas archaea *Methanomassiliicocales* was found to be more abundant in high efficiency (i.e., lower RFI) animals. Additionally, the authors suggested that more efficient animals harbour rumen microbiomes that better adapt to different environmental challenges and thus improve fermentation efficiency (Li et al., 2016). Paz et al. (2018) suggested that approximately 20% of the variation observed in gain:feed ratio could be explained by the microbiome in beef cattle, particularly by OTUs belonging to *Bacteroidales*, BS11, *Victivallaceae*, *Prevotellaceae*, *Fibrobacteraceae*, *Spirochaetae*, S24-7, *Paraprevotellaceae*, *Veillonellaceae* and *Lachnospiraceae*.

1.9 Association between the gastrointestinal microbiome and health traits in bovines

The gastrointestinal microbiome of bovine hosts not only greatly influences their feed conversion efficiency, due to their close association with the digestive processes, as presented in the previous sections, but also it is very closely associated with health traits.

Since birth, the establishment and development of a healthy and viable microbiota within the gastrointestinal tract of the ruminant is essential for the balance between health and disease. In ruminants, the microbiome profiles change rapidly and drastically throughout the first weeks of life, playing an important role in the development of immunity (Klein-Jöbstl et al., 2019). For example, in a very comprehensive review, Taschuk & Griebel (2012) suggested that the proper development of the mucosal immune system heavily depends on its exposure to a diverse gastrointestinal microbiota, which educates and trains the immune system. In a study including 60 Holstein cows,

Wang et al. (2021) showed that the content of ruminal VFAs in animals with clinical mastitis was significantly decreased in comparison to their subclinical or healthy counterparts, and that this had consequences such as reduced milk yield and quality (i.e., reduced fat and lactose content). Additionally, the rumen microbiota of animals with clinical mastitis also showed a decrease in commensals and an increase in potentially pathogenic microorganisms, which could be associated with intestinal or oral inflammation (Wang et al., 2021). In a review on the impact of nematode parasites on the gastrointestinal microbiome, Peachey et al. (2017) concatenates the results from several previous studies including hosts of different veterinary species, and suggests that microorganisms and metabolic markers involved in fibrolytic activities, and carbohydrate and protein transport and metabolism were altered as a consequence of parasitic infections. For example, Li et al. (2011) reported increased *Ethanoligenens* and decreased *Subdoligranulum* in the abomasum of dairy cattle infected with *Ostertagia ostertagi*.

1.10 Techniques to study the rumen microbiome

The study of the rumen microbiome often starts by the identification of its taxonomic composition, and the characterization of the abundances of each taxon. Since most of the rumen microbes are not culturable in laboratory conditions, other techniques of identification and quantification, such as those based on the 16S rRNA gene, and more recently, based on whole metagenomic sequencing of samples, have become extremely powerful to identify microbial profiles (Rappé and Giovannoni, 2003).

The most used technique for analysing microbiomes is the amplicon analyses of the 16S rRNA gene. This method is based on the PCR amplification of a segment of the 16S rRNA gene, which is then sequenced. The 16S rRNA gene is a highly conserved gene amongst prokaryotic organisms, however, it differs sufficiently between taxon, providing enough phylogenetic information as to allow for the identification of microorganisms at different taxonomic levels, from phylum to genus. In advantage to traditional clinical microbiology studies, the analysis of the 16S rRNA gene does not require the microorganisms to be

grown in culture plates. However, this method becomes inaccurate in the identification of microorganisms at the species level, mostly because of the great similarity of the 16S rRNA genes of different species. Furthermore, due to the variation between species of the number of 16S rRNA gene copies in the genomes, the accurate quantification of each taxon is impossible. Additionally, the amplification of each 16S rRNA gene may differ, due to variable affinity to the primers, which could introduce some bias to the characterization of the samples (Jo et al., 2016).

Whole genomic sequencing (WGS) techniques involve the random sequencing of DNA fragments in the sample, allowing for a comprehensive identification of microbial genes and organisms present in a complex sample. In advantage to 16S rRNA-based methods, WGS can be more accurate in the characterization of microbiome samples at the species level if good reference databases for the analysed community are available. Whereas 16S rRNA amplicon sequencing can be used to infer microbial gene content based on previous characterization of the identified OTUs, whole metagenome sequencing can be used to accurately resolve DNA fragments into microbial genes. Ranjan et al. (2016) compared the use of both methods in the characterization of a human stool microbiome sample and reported that WGS identified significantly more bacterial species per sample and identified even rare bacteria when both depth and coverage were adequate, with consequences on the sample's microbial diversity.

1.10.1 Statistical analyses of microbiome datasets

Microbiome datasets are compositional by nature, due to the way samples are processed. Sequencing instruments used in high-throughput sequencing can only deliver sequence reads up to the capacity of the instrument used, and therefore, the total read count is a fixed number, arbitrary for each run. The total number of reads in each sample is therefore not interesting, but rather the relationships between the components, which are often described as relative abundances (i.e., the abundance of a microbial taxon/gene divided by the total number of reads). The transformation of sequence read counts to relative

abundances is often performed to avoid dilution effects in the samples. However, the use of relative abundances directly in statistical analyses of microbiome datasets may lead to statistical pitfalls, for example from spurious correlations between parts, associated with subcompositional incoherence, i.e., results from statistics performed on the original dataset containing all observed variables differing from results from statistics performed in a subset of the original dataset (Gloor et al., 2017). The issue of subcompositional incoherence is discussed in the general discussion chapter of this thesis, along with other issues related to the compositionality of microbiome datasets, and some strategies (e.g., logratio transformations) to overcome these pitfalls.

1.11 Research objectives

This thesis addresses several research aims associated with the microbiome, firstly, the study of different methodologies for obtaining microbial communities, secondly, the estimation of associations of rumen microbiome profiles and bovine performance traits, thirdly, the evaluation of the stability of the rumen microbiome profiles and their usefulness to predict performance and methane traits based on longitudinal microbiome data, and fourthly, the influence of infection of animals with an abomasal nematode on the ruminal and caecal microbiome profiles. Therefore, the specific objectives are as follows:

1. To compare two popularly used bioinformatic pipelines, the MetaGenome Rapid Annotation using Subsystem Technology (MG-RAST) and Quantitative Insights Into Microbial Ecology 2 (QIIME2), based on 16S rRNA gene amplicons.

2. To understand the association between the bovine rumen metagenome and feed conversion efficiency traits, including FCR, RFI, ADG, and DFI.

3. To evaluate the longitudinal stability of the rumen microbiome composition throughout the finishing phase of beef cattle, and the ability

of the microbiome collected at different timepoints to predict performance and methane emissions traits.

4. To understand the impact of the parasitic abomasal nematode *Ostertagia ostertagi*, and of a vaccine against the parasite, on the rumen and caecum microbiome profiles of dairy cattle.

The aim of implementation of this research is the application of the results of these studies for breeding, dietary interventions, and vaccine development.

1.12 References

Accetto, T., and Avguštin, G. (2019). The diverse and extensive plant polysaccharide degradative apparatuses of the rumen and hindgut Prevotella species: A factor in their ubiquity? *Syst. Appl. Microbiol.* 42, 107–116. doi:10.1016/j.syapm.2018.10.001.

Ahring, B. K., Murali, N., and Srinivas, K. (2018). Fermentation of cellulose with a mixed microbial rumen culture with and without methanogenesis. *Ferment. Technol.* 07, 1–7. doi:10.4172/2167-7972.1000152.

Akin, D. E., and Borneman, W. S. (1990). Role of rumen fungi in fiber degradation. *J. Dairy Sci.* 73, 3023–3032. doi:10.3168/jds.S0022-0302(90)78989-8.

Anderson, C. L., Schneider, C. J., Erickson, G. E., Macdonald, J. C., and Fernando, S. C. (2016). Rumen bacterial communities can be acclimated faster to high concentrate diets than currently implemented feedlot programs. *J. Appl. Microbiol.* 120, 588–599. doi:10.1111/jam.13039.

Arthur, P. F., Archer, J. A., Johnston, D. J., Herd, R. M., Richardson, E. C., and Parnell, P. F. (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. doi:10.2527/2001.79112805x.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M. C. C., Charles, T., et al. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome* 8, 1–22. doi:10.1186/s40168-020-00875-0.

Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* 70, 567–590. doi:10.1152/physrev.1990.70.2.567.

Berry, D. P., and Crowley, J. J. (2013). Cell biology symposium: Genetics of feed efficiency in dairy and beef cattle. *J. Anim. Sci.* 91, 1594–1613. doi:10.2527/jas2012-5862.

Carberry, C. A., Kenny, D. A., Han, S., McCabe, M. S., and Waters, S. M. (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. doi:10.1128/AEM.07759-11.

Cheng, K. J., and Costerton, J. W. (1977). Ultrastructure of Butyrivibrio fibrisolvens: a gram positive bacterium? *J. Bacteriol.* 129, 1506–1512. doi:10.1128/jb.129.3.1506-1512.1977.

Department of Economic and Social Affairs Population Division - United Nations (2019). “World population prospects 2019 - Highlights,” in *World Population Prospects 2019* (New York: ONU), 5.

Difford, G. F., Plichta, D. R., Løvendahl, P., Lassen, J., Noel, S. J., Højberg, O., et al. (2018). Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genet.* 14, 1–22. doi:10.1371/journal.pgen.1007580.

Durunna, O. N., Mujibi, F. D. N., Goonewardene, L., Okine, E. K., Basarab, J. A., Wang, Z., et al. (2011). Feed efficiency differences and reranking in beef steers fed grower and finisher diets. *J. Anim. Sci.* 89, 158–167. doi:10.2527/jas.2009-2514.

FAO (2020). *Food security and nutrition in the world*.

Fernando, S. C., Purvis, H. T., Najar, F. Z., Sukharnikov, L. O., Krehbiel, C. R., Nagaraja, T. G., et al. (2010). Rumen microbial population dynamics during adaptation to a high-grain diet. *Appl. Environ. Microbiol.* 76, 7482–7490. doi:10.1128/AEM.00388-10.

Fliegerova, K., Kaerger, K., Kirk, P., and Voigt, K. (2015). “7. Rumen fungi,” in *Rumen Microbiology: From Evolution to Revolution* (India: Springer India), 97–112. doi:10.1007/978-81-322-2401-3.

Food and Agriculture Organization (FAO) (2019). Meat & Meat Products. *fao.org/ag/againfo/themes/en/meat/home.html*. Available at: fao.org/ag/againfo/themes/en/meat/home.html [Accessed March 24, 2021].

Food and Agriculture Organization (FAO) (2021). The role of ruminants in food security and livelihoods. *http://www.fao.org/in-action/enteric-methane/background/theroleofruminants/en/*. Available at: http://www.fao.org/in-action/enteric-methane/background/theroleofruminants/en/ [Accessed March 24, 2021].

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* 8, 1–6. doi:10.3389/fmicb.2017.02224.

Goad, D. W., Goad, C. L., and Nagaraja, T. G. (1998). Ruminal microbial and fermentative changes associated with experimentally induced subacute acidosis in steers. *J. Anim. Sci.* 76, 234–241. doi:10.2527/1998.761234x.

Godfray, H. C. J. C. J., Beddington, J. R. J. R., Crute, I. R. I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: The challenge of feeding 9 billion people. *Science (80-.).* 327, 812–818. doi:10.4337/9780857939388.

Gokarn, R. R., Eiteman, M. A., Martin, S. A., and Eriksson, K. E. L. (1997). Production of succinate from glucose, cellobiose, and various cellulosic materials by the ruminal anaerobic bacteria Fibrobacter succinogenes and Ruminococcus flavefaciens. *Appl. Biochem. Biotechnol. - Part A Enzym. Eng. Biotechnol.* 68, 69–80. doi:10.1007/bf02785981.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P. H., et al. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* 5, 14567. doi:10.1038/srep14567.

Holman, D. B., and Gzyl, K. E. (2019). A meta-analysis of the bovine gastrointestinal tract microbiota. *FEMS Microbiol. Ecol.* 95, 1–9. doi:10.1093/femsec/fiz072.

Jo, J. H., Kennedy, E. A., and Kong, H. H. (2016). Research techniques made simple: bacterial 16s ribosomal RNA gene sequencing in cutaneous research. *J. Invest. Dermatol.* 136, e23–e27. doi:10.1016/j.jid.2016.01.005.

Kaars Sijpesteijn, A. (1951). On Ruminococcus flavefaciens, a cellulose-decomposing bacterium from the rumen of sheep and cattle. *J. gen. Microbiol.* 5, 869–879.

Kenny, D. A., Fitzsimons, C., Waters, S. M., and McGee, M. (2018). Invited review: Improving feed efficiency of beef cattle – the current state of the art and future challenges. *Animal* 12, 1815–1826. doi:10.1017/s1751731118000976.

Klein-Jöbstl, D., Quijada, N. M., Dzieciol, M., Feldbacher, B., Wagner, M., Drillich, M., et al. (2019). Microbiota of newborn calves and their mothers reveals possible transfer routes for newborn calves' gastrointestinal microbiota. *PLoS One* 14, 1–18. doi:10.1371/journal.pone.0220554.

Krause, D. O., Denman, S. E., Mackie, R. I., Morrison, M., Rae, A. L., Attwood, G. T., et al. (2003). Opportunities to improve fiber degradation in the rumen: Microbiology, ecology, and genomics. *FEMS Microbiol. Rev.* 27, 663–693. doi:10.1016/S0168-6445(03)00072-X.

Kumar, S., Indugu, N., Vecchiarelli, B., and Pitta, D. W. (2015). Associative patterns among anaerobic fungi, methanogenic archaea, and bacterial communities in response to changes in diet and age in the rumen of dairy cows. *Front. Microbiol.* 6, 1–10. doi:10.3389/fmicb.2015.00781.

Li, F., Li, C., Chen, Y., Liu, J., Zhang, C., Irving, B., et al. (2019). Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome* 7, 1–17. doi:10.1186/s40168-019-0699-1.

Li, F., Zhou, M., Ominski, K., and Guan, L. L. (2016). Does the rumen microbiome play a role in feed efficiency of beef cattle? *J. Anim. Sci.* 94, 44–48. doi:10.2527/jas.2016-0524.

Li, R. W., Wu, S., Li, W., Huang, Y., and Gasbarre, L. C. (2011). Metagenome plasticity of the bovine abomasal microbiota in immune animals in response to ostertagia ostertagi infection. *PLoS One* 6. doi:10.1371/journal.pone.0024417.

Loor, J. J., Elolimy, A. A., and Mccann, J. C. (2016). Dietary impacts on rumen microbiota in beef and dairy production. *Anim. Front.*, 22–29. doi:10.2527/af.2016-0030.

Mackie, R. I., Gilchrist, F. M. C., Robberts, A. M., Hannah, P. E., and Schwartz, H. M. (1978). Microbiological and chemical changes in the rumen during the stepwise adaptation of sheep to high concentrate diets. *J. Agric. Sci.* 90, 241–254. doi:10.1017/S0021859600055313.

Martínez-Álvaro, M., Auffret, M. D., Duthie, C.-A., Dewhurst, R. J., Cleveland, M., Watson, M., et al. (2021). Bovine host genome acts on specific metabolism, communication and genetic processes of rumen microbes host-genomically linked to methane emissions. *Res. Sq. - Prepr.*, 0–37.

McCann, J. C., Wiley, L. M., Forbes, T. D., Rouquette, F. M., and Tedeschi, L. O. (2014). Relationship between the rumen microbiome and residual feed intake-efficiency of brahman bulls stocked on bermudagrass pastures. *PLoS One* 9, 1–6. doi:10.1371/journal.pone.0091864.

Meyer, A. M., Hess, B. W., Paisley, S. I., Du, M., Caton, J. S., and Du, M. (2014). Small intestinal growth measures are correlated with feed efficiency in market weight cattle, despite minimal effects of maternal nutrition during early to midgestation. *J. Anim. Sci.* 92, 3855–3867. doi:10.2527/jas.2014-7646.

Michalet-Doreau, B., Fernandez, I., and Fonty, G. (2002). A comparison of enzymatic and molecular approaches to characterize the cellulolytic microbial ecosystems of the rumen and the cecum. *J. Anim. Sci.* 80, 790–796. doi:10.2527/2002.803790x.

Montanholi, Y., Fontoura, A., Swanson, K., Coomber, B., Yamashiro, S., and Miller, S. (2013). Small intestine histomorphometry of beef cattle with divergent feed efficiency. *Acta Vet. Scand.* 55, 9. doi:10.1186/1751-0147-55-9.

Moran, J. (2005). “Chapter 5: How the rumen works,” in *Tropical dairy farming : feeding management for small holder dairy farmers in the humid tropics* (Collingwood: Csiro publishing), 41–49. doi:10.1016/0022-1694(84)90033-7.

Mukiibi, R., Vinsky, M., Keogh, K. A., Fitzsimmons, C., Stothard, P., Waters, S. M., et al. (2018). Transcriptome analyses reveal reduced hepatic lipid synthesis and accumulation in more feed efficient beef cattle. *Sci. Rep.* 8, 1–12. doi:10.1038/s41598-018-25605-3.

Nagaraja, T. G., and Titgemeyer, E. C. (2007). Ruminal acidosis in beef cattle: The current microbiological and nutritional outlook. *J. Dairy Sci.* 90, E17–E38. doi:10.3168/jds.2006-478.

Nozière, P., Besle, J. M., Martin, C., and Michalet-Doreau, B. (1996). Effect of barley supplement on microbial fibrolytic enzyme activities and cell wall degradation rate in the rumen. *J. Sci. Food Agric.* 72, 235–242. doi:10.1002/(SICI)1097-0010(199610)72:2<235::AID-JSFA647>3.0.CO;2-Y.

Nozière, P., Ortigues-Marty, I., Loncke, C., and Sauvant, D. (2010). Carbohydrate quantitative digestion and absorption in ruminants: from feed starch and fibre to nutrients available for tissues. *Animal* 4, 1057–1074. doi:10.1017/s1751731110000844.

Palevich, N., Kelly, W. J., Leahy, S. C., Denman, S., Altermann, E., Rakonjac, J., et al. (2020). Comparative genomics of rumen Butyrivibrio spp. uncovers a continuum of polysaccharide-degrading capabilities. *Appl. Environ. Microbiol.* 86, 1–19. doi:10.1128/AEM.01993-19.

Peachey, L. E., Jenkins, T. P., and Cantacessi, C. (2017). This gut ain't big enough for both of us. Or is it? Helminth–microbiota interactions in veterinary species. *Trends Parasitol.* 33, 619–632. doi:10.1016/j.pt.2017.04.004.

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* 469, 967–977. doi:10.1016/j.bbrc.2015.12.083.

Rappé, M. S., and Giovannoni, S. J. (2003). The uncultured microbial majority. *Annu. Rev. Microbiol.* 57, 369–394. doi:10.1146/annurev.micro.57.030502.090759.

Roehe, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., McKain, N., Ross, D. W., et al. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* 12, e1005846. doi:10.1371/journal.pgen.1005846.

Russell, J. B., and Hino, T. (1985). Regulation of Lactate Production in Streptococcus bovis: A Spiraling Effect That Contributes to Rumen Acidosis. *J. Dairy Sci.* 68, 1712–1721. doi:10.3168/jds.S0022-0302(85)81017-1.

Russell, J. B., Muck, R. E., and Weimer, P. J. (2009). Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. *FEMS Microbiol. Ecol.* 67, 183–197. doi:10.1111/j.1574-6941.2008.00633.x.

Tajima, K., Amoniv, R. I., Nagamine, T., Matsui, H., Nakamura, M., and Benno, Y. (2001). Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR. *Appl. Environ. Microbiol.* 67, 2766–2774. doi:10.1128/AEM.67.6.2766.

Taschuk, R., and Griebel, P. J. (2012). Commensal microbiome effects on mucosal immune system development in the ruminant gastrointestinal tract. *Anim. Health Res. Rev.* 13, 129–141. doi:10.1017/S1466252312000096.

Thornton, P. K. (2010). Livestock production: Recent trends, future prospects. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 2853–2867. doi:10.1098/rstb.2010.0134.

van Gylswyk, N. O., and Labuschagne, J. P. (1971). Relative efficiency of pure cultures of different species of cellulolytic rumen bacteria in solubilizing cellulose in vitro. *J. Gen. Microbiol.* 66, 109–113. doi:10.1099/00221287-66-1-109.

Wang, Y., Nan, X., Zhao, Y., Jiang, L., Wang, M., Wang, H., et al. (2021). Rumen microbiome structure and metabolites activity in dairy cows with clinical and subclinical mastitis. *J. Anim. Sci. Biotechnol.* 12, 1–21. doi:10.1186/s40104-020-00543-1.

Williams, C. L., Thomas, B. J., McEwan, N. R., Rees Stevens, P., Creevey, C. J., and Huws, S. A. (2020). Rumen protozoa play a significant role in fungal predation and plant carbohydrate breakdown. *Front. Microbiol.* 11, 1–14. doi:10.3389/fmicb.2020.00720.

Chapter 2 Taxonomic annotation of 16S rRNA sequences of pig intestinal samples using MG-RAST and QIIME2 generated different microbiota compositions

2.1 Introduction

An important step in the study of microbiome compositions is to assess the impact that the choice of bioinformatics pipeline used to resolve 16S rRNA gene amplicons into taxonomic groups has on the characterization of microbiota samples. In this chapter, I present our published journal article that compares the microbiota compositions of swine gastrointestinal and faecal samples obtained from using the bioinformatics pipelines MetaGenome Rapid Annotation using Subsystem Technology (MG-RAST), and Quantitative Insights Into Microbial Ecology 2 (QIIME2).

Journal of Microbiological Methods 186 (2021) 106235

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Taxonomic annotation of 16S rRNA sequences of pig intestinal samples using MG-RAST and QIIME2 generated different microbiota compositions

J. Lima [a,*], T. Manning [b], K.M. Rutherford [a], E.T. Baima [c], R.J. Dewhurst [a], P. Walsh [b], R. Roehe [a,*]

[a] Scotland's Rural College, Edinburgh, UK
[b] NSilico Life Science Ltd., Dublin, Ireland
[c] Zoetis Inc, Parsippany-, Troy Hills, United States

ARTICLE INFO

Keywords:
Taxonomic composition
MG-RAST
QIIME2
16S rRNA gene
Microbiota

ABSTRACT

Environmental microbiome studies rely on fast and accurate bioinformatics tools to characterize the taxonomic composition of samples based on the 16S rRNA gene. MetaGenome Rapid Annotation using Subsystem Technology (MG-RAST) and Quantitative Insights Into Microbial Ecology 2 (QIIME2) are two of the most popular tools available to perform this task. Their underlying algorithms differ in many aspects, and therefore the comparison of the pipelines provides insights into their best use and interpretation of the outcomes. Both of these bioinformatics tools are based on several specialized algorithms pipelined together, but whereas MG-RAST is a user-friendly webserver that clusters rRNA sequences based on their similarity to create Operational Taxonomic Units (OTU), QIIME2 employs DADA2 in the construction of Amplicon Sequence Variants (ASV) by applying an error model that considers the abundance of each sequence and its similarity to other sequences. Taxonomic compositions obtained from the analyses of amplicon sequences of DNA from swine intestinal gut and faecal microbiota samples using MG-RAST and QIIME2 were compared at domain-, phylum-, family- and genus-levels in terms of richness, relative abundance and diversity. We found significant differences between the microbiota profiles obtained from each pipeline. At domain level, bacteria were relatively more abundant using QIIME2 than MG-RAST; at phylum level, seven taxa were identified exclusively by QIIME2; at family level, samples processed in QIIME2 showed higher evenness and richness (assessed by Shannon and Simpson indices). The genus-level compositions obtained from each pipeline were used in partial least squares-discriminant analyses (PLS-DA) to discriminate between sample collection sites (caecum, colon and faeces). The results showed that different genera were found to be significant for the models, based on the Variable Importance in Projection, e.g. when using sequencing data processed by MG-RAST, the three most important genera were *Acetitomaculum*, *Ruminococcus* and *Methanosphaera*, whereas when data was processed using QIIME2, these were *Candidatus Methanomethylophilus*, *Sphaerochaeta* and *Anaerorhabdus*. Furthermore, the application of differential filtering procedures before the PLS-DA revealed higher accuracy when using non-restricted datasets obtained from MG-RAST, whereas datasets obtained from QIIME2 resulted in more accurate discrimination of sample collection sites after removing genera with low relative abundances (<1%) from the datasets. Our results highlight the differences in taxonomic compositions of samples obtained from the two separate pipelines, while underlining the impact on downstream analyses, such as biomarkers identification.

1. Introduction

The efficient and reproducible characterization of the microbial communities in a given sample (i.e. microbiota) is only as accurate as the bioinformatics tools applied to process the large rRNA amplicon sequencing datasets. Currently, the microbiota composition is widely explored using targeted 16S rRNA amplicons sequenced by a range of technologies which are fast and affordable in comparison to shotgun metagenomics (e.g. Andrade et al., 2020; Koringa et al., 2019; Li et al., 2016). Several pipelines are available to perform quality checks and taxonomic annotation: Metagenome Analyzer (MEGAN) is a computer program that uses NCBI annotation to annotate reads according to their conservation level (Huson et al., 2007); MOTHUR is a pipeline that allows the user to trim, screen and align sequences and includes tools to

* Corresponding authors.
E-mail addresses: joana.lima@sruc.ac.uk (J. Lima), rainer.roehe@sruc.ac.uk (R. Roehe).

https://doi.org/10.1016/j.mimet.2021.106235
Received 1 March 2021; Received in revised form 4 May 2021; Accepted 6 May 2021
Available online 8 May 2021

evaluate diversity parameters, such as alpha- and beta-diversity, and visualization tools such as Venn diagrams and heat maps (Schloss et al., 2009); Metagenomics Rapid Annotation using Subsystem Technology (MG-RAST, Meyer et al., 2008) is a fully automated pipeline that employs similarity-based binning of rRNA sequences into Operational Taxonomic Units (OTUs), followed by the comparison of each OTU representative against the M5rna database, using the Blast-Like Alignment Tool (BLAT, Glass et al., 2010); Quantitative Insights into Microbial Ecology (QIIME2, Bolyen et al., 2018) is an open-source software for the analysis of microbiomes that employs the Divisive Amplicon Denoising Algorithm package (DADA2, Callahan et al., 2016) in an Amplicon Variant Sequence (ASV)-based binning of sequences.

In previous studies, D'Argenio et al. (2014) and Plummer et al. (2015) compared MG-RAST to QIIME (identifying OTUs), a precursor of QIIME2 (Identifying ASVs), and both found that MG-RAST consistently reported a significantly higher number of unclassified sequences than QIIME. However, whereas D'Argenio et al. (2014) concluded that QIIME provided more accurate results than MG-RAST, Plummer et al. (2015) concluded that the tools generated similar results. QIIME and QIIME2 are two pipelines that include external tools to perform some specific tasks, i.e. QIIME2 is not a true update of QIIME, as they substantially differ in the set of tools and algorithms they employ; e.g. whereas QIIME is an OTU-based pipeline, QIIME2 is ASV-based. Kaszubinski et al. (2019) compared MG-RAST, MOTHUR and QIIME2, based only on the phylum- and family-level compositions, after a rarefaction procedure and filtering out the OTUs with mean relative abundance lower than 1%, and suggested that QIIME2 was the most appropriate pipeline, mostly due to decreased abundance of unclassified sequences, differentially abundant taxa and increased alpha- and beta-diversity in comparison to MG-RAST and MOTHUR. Although MOTHUR is widely used to analyse community sequence data, it was not included in the present study because it has been previously found to produce the highest percentage of unclassified reads when compared to MG-RAST and QIIME, at the phylum- and family-level (Kaszubinski et al., 2019) and at the genus-level (Plummer et al., 2015) compositions. Furthermore, MOTHUR had the most false positives and lowest concordance to the microbiota taxonomic reference dataset (Kaszubinski et al., 2019).

The aim of the present research was to compare the taxonomic compositions resulting from the application of the two pipelines, MG-RAST and QIIME2, when sequences are aligned against the SILVA database (Pruesse et al., 2007). These pipelines were selected for comparison because they are self-contained and have fundamentally different underlying algorithms. Both tools are among the most popular freely available software used to obtain taxonomic composition of samples from 16S rRNA amplicon sequence files provided by Illumina procedures (i.e. raw sequence reads in fastq format), and both allow the use of the SILVA database (Quast et al., 2013) in the identification of OTUs/ASVs. Additionally, we investigated the use of microbiota profiles obtained using MG-RAST and QIIME2 in the discrimination of sample collection sites (caecum, colon and faeces), and assessed the potential consequences of using distinct tools for microbiota characterization of samples. The impact of different filtering and data cleaning processes on the microbiota composition were also evaluated.

2. Materials and methods

2.1. Ethical statement

The porcine trial was conducted at the Pig Research Centre of Scotland's Rural College (SRUC, 6 miles south of Edinburgh, UK). The experiment was approved by SRUC's Animal Welfare and Ethical Approval Body and was conducted in accordance with the requirements of the UK Animals (Scientific Procedures) Act 1986.

2.2. Bioinformatics pipelines

The 16S rRNA amplicon reads obtained from an Illumina MiSeq System (Edinburgh Genomics, UK) were analysed using two pipelines: MG-RAST (v. 4.0.3) and QIIME2 (v. 2019.1). Pre-processing of samples, such as quality-based trimming and/or filtering and chimera detection and removal, were performed only if the necessary tools were provided within the pipelines because the aim of this study was to apply MG-RAST and QIIME2 as stand-alone tools from start (16S rRNA reads obtained from Illumina in fastQ files) to finish. For both tools, the SILVA reference database (SSU, release 132) was used in the taxonomic annotation of reads.

2.2.1. QIIME2

2.2.1.1. Data hygiene. High-throughput sequencing techniques such as the one employed in Illumina sequencers exhibit a steep, exponential increase in error rates along the read length. Illumina results include both the nucleotide sequence of the reads and a quality score (Q-score) associated to each nucleotide in each read. The QIIME2 pipeline uses the Q-scores for the quality-score-based trimming and filtering procedure, by randomly selecting a subset of reads per base position and calculating a boxplot of the corresponding Q-scores. These results are then provided to the user, who makes the decision regarding trimming. Forward and reverse reads were trimmed at 153 and 157 bases, respectively. No external tool for chimera detection and removal was actively included in our protocol using either pipeline, but DADA2 (incorporated in QIIME2) defaults the action regarding chimeras to "consensus" (i.e. "Chimeras are detected in samples individually, and sequences found chimeric in a sufficient fraction of samples are removed." (QIIME2 Development Team, 2020)). In this work, we used a typical or best-practices approach to QIIME2, following official tutorials.

2.2.1.2. Feature annotation. QIIME2 employs DADA2 in the identification of ASVs. DADA2 is an open-source software package for modelling and correcting Illumina-sequenced amplicon errors (Callahan et al., 2016). This algorithm implements a quality-aware model that works by first grouping all amplicon reads with the same sequence into unique-sequence sets, keeping record of the abundance and consensus quality profile of each of these sets. All unique-sequence sets are grouped into one single partition and the most abundant one (the one with the highest amount of copies) becomes the centre of the partition. The similarity of each unique-sequence set to the centre set is calculated and the one with highest dissimilarity from the centre set is identified to become the centre of a new partition. All unique-sequence sets are then re-distributed through the two partitions, according to their similarity to the centre sets. This process continues iteratively until the division of sequences into partitions is consistent with the error model creating ASVs (Callahan et al., 2016). DADA2 algorithm is mainly based on two criteria; the abundance of each amplicon sequence (if a sequence is highly abundant, it is most likely a product of true variation than a product of errors introduced during the sequencing procedure) and the pairwise similarity between sequences (i.e. error rates). Then, a classifier is used in the identification of each ASV. In this study, a Bayesian Naïve classifier was pre-trained on the SILVA database ("silva-132-99-515-806-nb-classifier.qza", *https://docs.qiime2.org/2019.4/data-resources/*) and then used for the taxonomic annotation of our samples. One table was created for each taxonomic level (Domain, Phylum, Class, Order, Family and Genus). In each table, sequences that could not be allocated to the corresponding taxonomic level (but were allocated to any higher level) were accumulated into a new category named "Unidentified".

2.2.2. MG-RAST

2.2.2.1. Data hygiene. Forward and reverse read files were uploaded

into MG-RAST, which performs an automated quality control step based on an md5 checksum, to identify any issues with the sequencing run, such as corrupt files. No issues were detected in our files. At the same time, it provides statistical information about the reads present in each file, such as base pair count, sequence count, sequence length and GC-content. MG-RAST includes options for demultiplexing (necessary when reads from multiple samples are mixed in one file, which was not the case in our study) and joining paired-ends. In this pipeline, the pairing must be performed before any quality-based trimming and filtering of reads. MG-RAST incorporates the SolexaQA software package (Cox et al., 2010), which provides a rapid assessment of read quality for data generated using Illumina sequencing; the user can select a threshold quality for trimming (the default value is Q ~ 13 or $P = 0.05$). In our study, the threshold applied was Q ~ 25, as to warrant base call accuracy not to be lower than 99.5%. Although in other pipelines (such as QIIME2) the quality-based trimming is an important procedure, with consequences for downstream analyses, in MG-RAST this is not as critical because MG-RAST offers the user the possibility of filtering data after the taxonomic annotation process, according to criteria such as the minimum identity cut-off, the minimum alignment length, the minimum abundance threshold and the maximum e-value. No procedure for chimera detection and removal are provided within MG-RAST. We did not use any external software for this purpose, because the goal was to exclusively use the tools provided in each pipeline. The quality control procedures in MG-RAST also include a screening stage that uses Bowtie (Langmead et al., 2009) to discard sequences that are near-exact matches to the host genomes of a selected organism, in our case *Sus scrofa* (i.e. the wild boar).

2.2.2.2. Feature annotation. Following the quality control, MG-RAST performs a rRNA extraction at 70% identity using VSEARCH (Rognes et al., 2016) against a 90% identity clustered reduced version of SILVA, Greengenes (DeSantis et al., 2006) and RDP (Cole et al., 2003) databases (M5RNA_90). The rRNA reads identified are then clustered at 97% of identity using cd-hit (Li and Godzik, 2006) and the longest sequence is picked as the cluster representative for comparison against the M5rna database (non-redundant database that includes SILVA, Greengenes and RDP, Wilke et al., 2012) using the BLAST-like Alignment Tool (BLAT; Kent, 2002). In this study, we opted to use the "SILVA SSU" option, which will provide the SILVA annotation of the features, based on the M5rna database. After the feature annotation, MG-RAST provides the possibility to manipulate several parameters in real time using the user interface (UI), including the minimum identity cut-off and the minimum alignment length which were in the present study set to 80% and 100, respectively. Additionally, the user can also define the minimum abundance threshold and the maximum e-value (or expect value, i.e. the number of hits one can expect to be by chance in the used database), which were left at the default value.

2.3. Data used for methodology comparison

For the comparisons of MG-RAST and QIIME2, 188 samples of the gastrointestinal microbiota were obtained from 38 intact male swine, which were progeny of crosses of Hampshire boars and Large White × Landrace crossbred sows (38 samples from caecum and colon each, and 112 faecal samples). Faeces samples were collected at 3 time points (start, at the end of the second week and at the end of the fourth week). For each faecal sample, about 5 g of homogenized matter was placed in 30 ml universal containers (Alphalabs, UK) filled with 4 ml RNALater (Sigma-Aldrich, UK) prior to being snap frozen and stored at −80 °C. At the end of the trial, pigs were sedated and euthanized prior to dissection and tissue collection. Post-mortem, intestinal luminal contents were collected from the caecum and colon using Universal 30 ml tubes, whereas mucosal cell wall samples were captured by scraping and transferred into Nunc 4.5 ml cryotubes. Tubes were filled with 4 ml or 3 ml of RNALater, respectively. Total DNA was extracted from intestinal content and faeces samples following an adapted protocol of Yu and Morrison (2004) combining chemical lysis and bead beating followed by purification on column using the QIASymphony with the Qiagen Midi kit and applying the blood sample extraction method with the FIX option used to collect all the supernatant. DNA was finally eluted in 400 μl of EB (Qiagen, UK) and an aliquot of 200 μl was directly stored at −20 °C whilst a second one was retained for further analysis. The amount of DNA extracted was quantified by Qubit fluorimetric quantitation for dsDNA (ThermoFisher, UK). An adapted protocol based on the 16S Metagenomic Sequencing Library Preparation for the Illumina MiSeq System (Illumina, UK) was applied for total DNA extracted from caecum, colon and faeces samples. The V4 region of the 16S rRNA gene was amplified specifically using primers 515F and 806R. Two 16S libraries were composed of 95 and 93 amplicon samples purified on magnetic beads using the ProNex Chemistry (Promega, WI, USA) and quantified using Qubit assay prior to being pooled in two different tubes. An aliquot of 10 ng/μl in 15 μl per library was sent to Edinburgh Genomics (Scotland, UK) for Illumina sequencing using MiSeq v2 250PE and providing a yield of at least 11 M + 11 M reads per run, resulting in a total of 376 fasta.gz files (forward and reverse read files for each of the 188 samples).

2.4. Statistical analyses

The relative abundances of each group at domain level, of unclassified and unidentified sequences at phylum-level and at genus-level were calculated from the taxonomic composition reported by MG-RAST and QIIME2 and compared using two-sided paired *t*-tests, considering the different variances of the samples.

In the domain-level analysis and in all analyses pertaining to the percentage of unclassified and unidentified data, all reported groups were considered, including the sequences classified as Eukaryota and/or viruses. However, for the phylum-, family- and genus-level analyses, these groups were removed from the databases. To account for multiple testing, presented *P*-values were adjusted by applying the Bonferroni correction.

Venn diagrams were used to identify taxa reported exclusively by either QIIME2 or MG-RAST and by both, and were applied to the phyla-, family- and genus-level tables (Oliveros, 2007).

Principal Component Analysis (PCA) and plots were produced using devtools (Hadley et al., 2019) and ggbiplot (Vu, 2011) packages in R Studio (v. 1.1.453). PCA was carried out using the relative abundances of the 56 families reported by both MG-RAST and QIIME2, after removing four outliers (corresponding to two samples whose taxonomic compositions generated by either pipeline diverged from the mean by more than 4 standard deviations). Additionally, the family compositions obtained from MG-RAST and QIIME2 were compared in a permutation multivariate analysis of variance (PERMANOVA), using a Bray-Curtis distances matrix (Anderson, 2001; Mcardle and Anderson, 2001). This analysis was carried out using the adonis() function of the vegan package (Oksanen et al., 2019) in R Studio (v. 1.1.453).

tThe correlation of the absolute counts of each genus identified by both pipelines was calculated, and significance was assessed through Bonferroni adjusted *P*-value (P-value <0.05 means significant correlation).

Partial Least Squares Discriminant Analyses (PLS-DA), calculated with the 'mixOmics' package (Le Cao et al., 2020) in R Studio (v. 1.1.453), was used to address differences in the microbiota profiles of 114 samples collected from different body sites (caecum, colon and faeces) at the end of the trial. These analyses were performed (in relative abundances) after subjecting the genus-level datasets generated by MG-RAST and QIIME2 to four different filtering procedures: data scenario A included all genera reported by each pipeline; data scenario B included genera reported by each pipeline with average relative abundance greater than 1%; data scenario C included genera identified by both pipelines; data scenario D included all taxa identified by both pipelines with a minimum average relative abundance of 1% (Table 1). Relative abundances were calculated

Table 1
Filtering criteria applied to genus-level compositions to create 4 data scenarios for comparisons.

Data scenario	Criteria		Total number of genera used
	Minimum average relative abundance	Genera identified in either or both pipelines	
A	Not applied	Either	MG-RAST: 225 genera QIIME2: 159 genera
B	Applied	Either	MG-RAST: 14 genera QIIME2: 22 genera
C	Not applied	Both	Both: 86 genera
D	Applied	Both	Both: 10 genera

within each data scenario. Considering the compositional nature of microbiota datasets, further PLS-DA analyses were performed on data scenarios A and B using the MG-RAST and QIIME2 datasets transformed by the additive logratio methodology (Greenacre, 2018). The denominators (*Acidaminococcus* in MGRAST and *Subdoligranulum* in QIIME2) were identified based on their high prevalence, low variance, and high Procrustes correlations with the full log-ratio space.

The co-abundances of microbial taxa in each of the data scenarios were explored in networks analyses using Graphia Professional software (Kajeka Ltd., Edinburgh; Freeman et al., 2007), in which nodes represent samples and edges represent a correlation value above $R = 0.95$. Clustering was performed using the Markov clustering method (MCL) available in Graphia Professional using the default settings (inflation, pre-inflation, and scheme values of 6). Each cluster was evaluated for enrichment of samples according to their collection sites (caecum, colon and faeces) and pipeline (MG-RAST and QIIME2), whereby enrichment significance was identified at P-value <0.05.

2.4.1. Diversity measures

Several diversity measurements were used to compare the family composition of the samples as obtained from MG-RAST and QIIME2 pipelines.

2.4.1.1. Alpha-diversity. Observed richness and gamma-diversity (S_{obs} and γ, respectively) are both richness measurements. S_{obs} corresponds to the number of different taxa in each individual sample. For a group of samples, S_{obs} corresponds to the average number of taxa per sample, whereas γ corresponds to the total number of different taxa in a collection of samples.

The *Chao*1 index (Chao, 1987) refers to the richness estimated for each sample (including all taxa measured by S_{obs} and taxa that were presumably not sampled). The *Chao1* index was calculated using the fossil package (Vavrek, 2015) in R Studio (v. 1.1.453).

Shannon and Simpson indices (H' and D', respectively) are abundance-based measures of diversity. The results presented here refer to the adjusted Shannon and adjusted Simpson indices (*H'adj* and *D'adj*, respectively), which correspond to the ratios of H' and D' by the maximum H' and maximum D' possible, respectively (Veech, 2017). Note that H', *H'adj*, D' and *D'adj* increase with diversity (Veech, 2017).

2.4.1.2. Beta-diversity. Multiple beta-diversity measures can be used to infer diversity between samples.

The additive and multiplicative partitioning of the gamma-diversity are the most direct ways of calculating beta-diversity; beta-additive (β_A) corresponds to the taxa richness typically absent from a randomly selected sample whereas beta-multiplicative (β_M) refers to the number of unique samples (i.e. with no taxa in common with any other sample) theoretically found in a group of samples.

The Bray-Curtis index (C_{BC}) makes use of abundance data to calculate diversity between samples in a pairwise manner, which results in a vector of distances to every other sample; the average value corresponds to the C_{BC} presented.

β_A, β_M and C_{BC} were used here to address dissimilarity of samples within their originating pipeline (Veech, 2017).

3. Results

3.1. Domain

The taxonomic composition of 188 samples computed in MG-RAST and QIIME2 resulted in totals of 20,760,260 and 14,576,856 hits, respectively (Table 2). The average relative abundances of Archaea were non-significantly different, at 0.47 ± 0.88% and 0.42 ± 0.57% using MG-RAST and QIIME2, respectively, whereas the differences in relative abundances of Unclassified, Bacteria and Eukaryota were significant (Table 2). The main reason for these differences was the misclassification of 16S rRNA amplicon sequences into Eukaryota and Viruses, and a relatively higher percentage of unclassified sequences in MG-RAST.

The Archaea:Bacteria ratio of each sample was calculated using absolute counts of hits obtained from each pipeline, and then compared in a two-sided paired *t*-test, which revealed a significant difference between the pipelines, resulting mostly from the differential abundance of bacteria.

3.2. Phylum

The relative abundances of unclassified and unidentified sequences were significantly higher in the phylum-level composition of samples obtained from MG-RAST than those from QIIME2 (0.017% and 3.483%, respectively, P-value<0.001). All 15 phyla reported by MG-RAST were also identified by QIIME2, whereas 7 were exclusively identified by QIIME2 (Fig. 1). The overall abundances of the 15 phyla found in common between both pipelines accounted for 100% and 98.8% of the QIIME2 and MG-RAST total hits, respectively, indicating that taxa identified exclusively by QIIME2 only accounted for a small part of the overall microbiota. Despite this, some of the phyla exclusively reported by QIIME2 had higher average relative abundance (over all animals) than phyla identified by both pipelines, e.g. *Euryarchaeota* was identified by both pipelines and accounted for 0.4% of QIIME2 hits, whereas *Kiritimatiellaeota*, exclusively identified by QIIME2, accounted for 0.7%.

The relative abundances of 13 of the 15 phyla were shown to be significantly different between the pipelines (P-value<0.05, Table 3).

Table 2
Mean relative abundances (%) of each domain as reported by MG-RAST and QIIME2 pipelines. This table summarises the total number of sequences identified, the averages and standard deviations of the relative abundances of each domain and whether significant differences occur depending on the used pipeline.

Taxonomic level	Pipeline		P-value
	MG-RAST	QIIME2	
Archaea	0.47 ± 0.88	0.42 ± 0.57	0.1681
Bacteria	96.02 ± 5.64	99.57 ± 0.57	2.20E-16
Eukaryota	3.36 ± 5.20	3.77E-04 ± 1.91E-03	5.73E-16
Viruses	2.94E-05 ± 2.46E-04	–	–
Unclassified	0.15 ± 0.48	5.24E-03 ± 1.56E-02	5.85E-05
A:B ratio	5.19E-03 ± 1.01E-02	4.28E-03 ± 0.06	0.0202
Total sequences	20,760,260	14,576,856	–

A:B ratio indicates Archaea:Bacteria ratio. P-values were obtained from two-sided paired t-tests, assuming different variances.

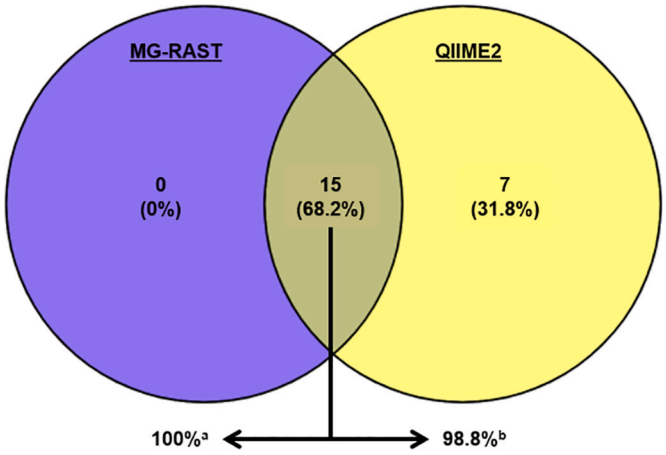


Fig. 1. Commonly and differently identified phyla reported by MG-RAST and QIIME2. The numbers in brackets represent percentage of taxa in each area. The 15 phyla shared by the pipelines correspond to (a) 100% of the sequence hits reported by MG-RAST and (b) 98.8% of the sequence hits reported by QIIME2.

Table 3
Differences between the relative abundances of the phyla identified by MG-RAST and QIIME2 pipelines. 'Mean of differences' refers to the average difference between relative abundances reported by the pipelines and the '*P*-value Bonferroni' indicates the significance of those differences. These relative abundances were calculated considering only the taxa obtained with both pipelines, in order to allow for a balanced comparison.

Phylum	Average relative abundances (%)		Mean of differences	P-value Bonferroni
	MG-RAST	QIIME2		
Firmicutes	62.37 ± 6.29	47.98 ± 6.47	14.4 ± 4.99	1.34E-91
Bacteroidetes	35.32 ± 7.030	41.88 ± 5.96	6.79 ± 4.06	3.17E-48
Actinobacteria	1.23 ± 0.97	2.40 ± 0.89	1.23 ± 0.5	3.63E-61
Euryarchaeota	0.53 ± 1.01	0.43 ± 0.59	0.2 ± 0.5	0.038481
Proteobacteria	0.51 ± 0.53	2.98 ± 1.86	2.47 ± 1.89	2.2E-41
Tenericutes	1.27E-02 ± 3.48E-02	0.87 ± 0.55	0.86 ± 0.54	1.44E-51
Spirochaetes	1.05E-02 ± 7.43E-02	1.68 ± 2.50	1.67 ± 2.47	3.35E-16
Fusobacteria	9.28E-03 ± 7.04E-02	5.88E-03 ± 4.14E-02	0.004 ± 0.03	0.118873[a]
Fibrobacteres	3.92E-03 ± 2.19E-02	0.25 ± 0.38	0.25 ± 0.38	1.56E-15
Chlamydiae	7.91E-04 ± 2.45E-03	6.10E-04 ± 2.73E-03	0.0005 ± 0.001	0.118873[a]
Lentisphaerae	6.94E-04 ± 4.93E-03	0.13 ± 0.17	0.13 ± 0.17	4.44E-19
Cyanobacteria	1.84E-04 ± 1.07E-03	1.34 ± 1.11	1.34 ± 1.11	1.07E-37
Synergistetes	2.13E-05 ± 2.92E-04	1.30E-02 ± 1.34E-02	0.01 ± 0.01	9.43E-28
Verrucomicrobia	1.35E-05 ± 1.31E-04	4.90E-03 ± 7.39E-03	0.005 ± 0.01	1.06E-15
Elusimicrobia	7.96E-06 ± 7.77E-05	3.74E-02 ± 6.74E-02	0.04 ± 0.07	5.1E-12

[a] indicates non-significant differences ($p \geq 0.05$). *P*-values were obtained from two-sided paired *t*-tests assuming different variances and corrected by the Bonferroni method.

The highly abundant taxa showed differences in relative abundances as result of the pipeline used, e.g. *Firmicutes* was significantly more abundant using MG-RAST (62.3%) than QIIME2 (48.0%), and *Bacteroidetes* was significantly more abundant using QIIME2 (35.3%), than MG-RAST (41.9%).

To highlight the differences in taxa with lower abundances, the logarithms of the relative abundances were used in a mirrored bar chart (Fig. 2) in which the longer bars correspond to less abundant taxa e.g. SAR in QIIME2 and *Elusimicrobia* in MG-RAST.

The analyses of the phyla prevalence in the samples revealed that *Actinobacteria*, *Bacteroidetes*, *Firmicutes* and *Proteobacteria* were identified in all samples, independently of the pipeline used (Table 4). Additionally, *Cyanobacteria* and *Tenericutes* were present in all compositions reported by QIIME2, but only in 5% and 53% when applying MG-RAST, respectively. Six out of 15 phyla and 17 out of 22 phyla were detected in at least half of the samples by MG-RAST and QIIME2, respectively, and these summed up to average relative abundances of 99.97% and 99.95%, respectively. Although seven phyla were exclusively reported by QIIME2, five of these were present in at least 50% of the samples.

3.3. Family

The family-level composition of samples obtained from the different pipelines (after removal of unclassified, unidentified and Eukaryote groups) was investigated in a Venn diagram (Fig. 3). A large number of taxa were found to be exclusively identified by one of the pipelines: 47 and 25 for MG-RAST and QIIME2, respectively.

The 10 most abundant families from each pipeline were ranked by decreasing average relative abundance in Table 5. *Prevotellaceae* and *Ruminococcaceae* were the 1st and 2nd most abundant taxa in both pipelines. An additional five taxa (*Veillonellaceae*, *Clostridiaceae*, *Lachnospiraceae*, *Erysipelotrichaceae* and *Lactobacillaceae*) were in the top 10 most abundant taxa of both pipelines, although not in the same rank e.g. *Veillonellaceae* was the 3rd most abundant in MG-RAST but only the 5th in QIIME2. *Eubacteriaceae*, *Acidaminococcaceae* and *Coriobacteriaceae* were among the 10 most abundant families as reported by MG-RAST, but were identified at lower abundances by QIIME2. Among the 10 most abundant families identified by QIIME2, *Muribaculaceae* was not detected by MG-RAST, and both *Peptostreptococcaceae* and *Rikenellaceae* were identified, but at lower relative abundances.

3.3.1. Taxonomic diversity

The taxonomic diversity was assessed and compared using different indices (Table 6). The family compositions of each sample obtained from QIIME2 showed on average greater richness than their MG-RAST counterparts (S_{obs} = 38.95 and 30, respectively). In contrast, sample compositions obtained from QIIME2 had lower overall richness than MG-RAST (γ = 81 and 103, respectively). *Chao*1 (which takes into account both the observed and unobserved richness of samples) were significantly lower when samples were characterized using MG-RAST. However, there was a large difference between *Chao*1 and S_{obs} when using MG-RAST, indicating a higher estimated number of unsampled families. In contrast, *Chao*1 estimated for samples compositions obtained from QIIME2 had the same values as S_{obs}. This could lead to the conclusion that QIIME2 provided us the absolutely complete characterization of the targeted environments (caecum, colon and faeces), i.e. not one family that exists in the sampled environments has been left unsampled, which is virtually impossible. Instead, this is because DADA2 (incorporated in QIIME2) does not call singletons, due to the difficulty in robustly distinguishing real singletons from singleton errors (Callahan et al., 2016).

The adjusted Shannon and Simpson indices showed that samples had, on average, significantly lower evenness and higher dominance (as indicated by lower *H'adj* and lower *D'adj*) when the taxonomic compositions were obtained from MG-RAST. Dissimilarity between samples was assessed through beta partitioning and Bray–Curtis dissimilarity index. The β_A was higher when MG-RAST was applied, indicating that the composition of each sample was more similar to the average composition of the collection of samples when these were processed in QIIME2. The β_M was also higher when using MG-RAST in comparison to QIIME2, suggesting that microbiota profiles obtained from the latter pipeline were more similar to each other than those obtained from the

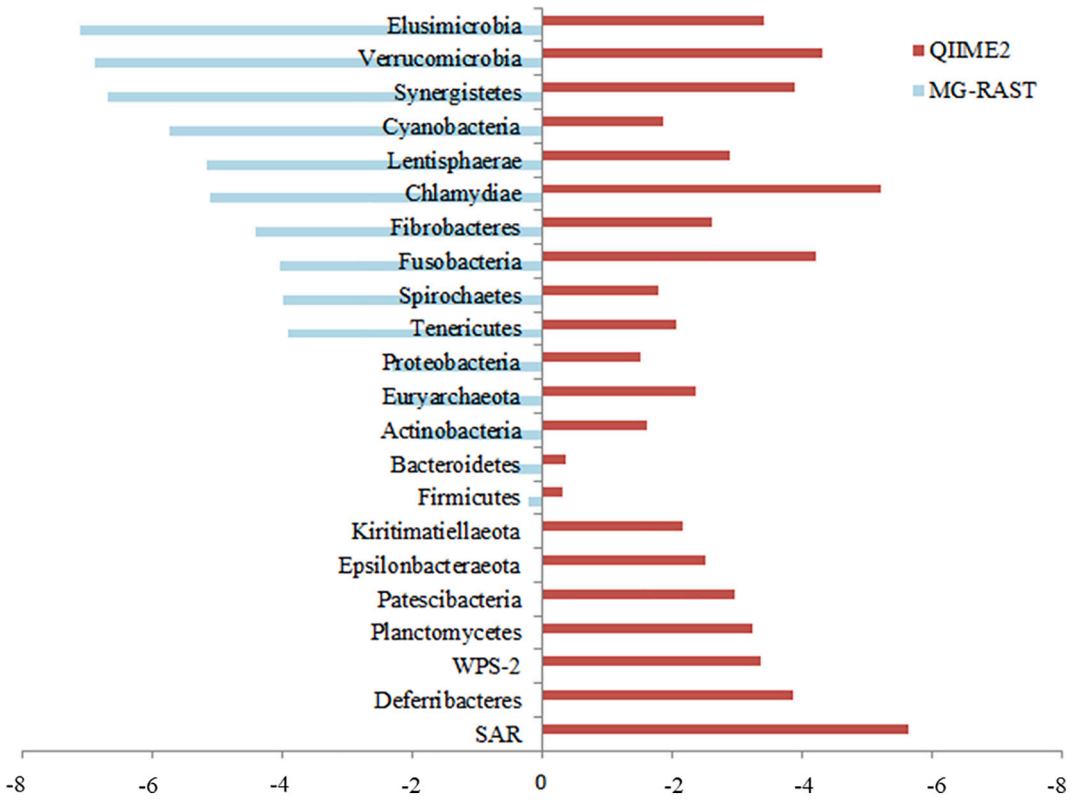


Fig. 2. Log-transformed relative abundances of the 22 phyla obtained using MG-RAST and/or QIIME2. Note that shorter bars correspond to higher abundances.

Table 4
Prevalence of each phylum in 188 samples.

Phyla	Prevalence in QIIME2 (%)	Prevalence in MG-RAST (%)
Actinobacteria	100	100
Bacteroidetes	100	100
Cyanobacteria	100	4.79
Firmicutes	100	100
Proteobacteria	100	100
Tenericutes	100	53.19
Spirochaetes	98.94	23.40
Fibrobacteres	93.09	4.26
Euryarchaeota	86.17	95.21
Lentisphaerae	76.60	5.32
Synergistetes	75.53	0.53
Elusimicrobia	52.66	1.06
Verrucomicrobia	42.02	1.06
Fusobacteria	19.15	35.11
Chlamydiae	7.45	24.47
Epsilonbacteraeota	98.94	n/a
Patescibacteria	86.70	n/a
Kiritimatiellaeota	82.98	n/a
Planctomycetes	65.43	n/a
Deferribacteres	50.53	n/a
WPS-2	41.49	n/a
SAR	2.13	n/a

n/a indicates non-applicable, these phyla were identified exclusively by QIIME2.

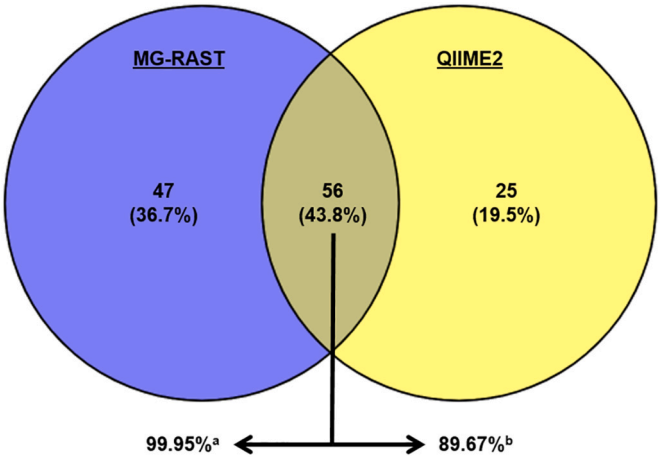


Fig. 3. Commonly and differently identified families using MG-RAST and QIIME2. The numbers in brackets represent the proportion of families in each area. A total of 56 families were reported by both pipelines and enclosed (a) 99.95% of the sequence hits in MG-RAST and (b) 89.67% of the sequence hits in QIIME2.

former pipeline. Similarly, C_{BC} was significantly higher in MG-RAST than in QIIME2, indicating that family-level composition of samples obtained from MG-RAST were more dissimilar between themselves than when using QIIME2; C_{BC} differs from beta partitioning measures because it is calculated as the average of pairwise differences, rather than as a collective group.

The PCA plot in Fig. 4 showed that the microbiota profiles (including the 56 families detected by both MG-RAST and QIIME2) formed different clusters, depending on the pipeline from which they were obtained. Additionally, PERMANOVA analysis showed significant differences (P-value <0.01) in the microbiota communities, with higher relative abundances of *Eubacteriaceae*, *Clostridiaceae*, *Veillonellaceae* and *Acidaminococcaceae* when using MG-RAST and higher relative abundances of *Succinivibrionaceae*, *Peptostreptococcaceae*, *Ruminococcaceae*, *Rikenellaceae* and *Lachnospiraceae* in results obtained from QIIME2.

3.4. Genus

The similarity of genus-level composition of samples between the use of MG-RAST and QIIME2 was explored in the Venn diagram in Fig. 5.

Table 5
The 10 most abundant families reported by each pipeline and corresponding overall samples average relative abundances.

MG-RAST			QIIME2		
Rank	Family	Average Relative Abundance (%)	Rank	Family	Average Relative Abundance (%)
1	*Prevotellaceae*[a]	36.54 ± 1.74E-01	1	*Prevotellaceae*[a]	33.40 ± 9.21
2	*Ruminococcaceae*[a]	15.91 ± 1.85E-02	2	*Ruminococcaceae*[a]	15.81 ± 3.85
3	*Veillonellaceae*[b]	10.83 ± 1.80E-04	3	*Lachnospiraceae*[b]	10.82 ± 2.53
4	*Clostridiaceae*[b]	9.11 ± 7.48E-03	4	*Muribaculaceae*	6.84 ± 4.35
5	*Eubacteriaceae*	6.41 ± 1.86E-04	5	*Veillonellaceae*[b]	6.31 ± 3.20
6	*Lachnospiraceae*[b]	5.04 ± 5.79E-04	6	*Erysipelotrichaceae*[b]	5.35 ± 1.85
7	*Erysipelotrichaceae*[b]	4.98 ± 4.13E-02	7	*Lactobacillaceae*[b]	3.39 ± 3.84
8	*Lactobacillaceae*[b]	4.66 ± 2.57	8	*Clostridiaceae*[b]	2.48 ± 4.40
9	*Acidaminococcaceae*	3.84 ± 1.32E-03	9	*Peptostreptococcaceae*	1.84 ± 3.86
10	*Coriobacteriaceae*	1.01 ± 1.15E-04	10	*Rikenellaceae*	1.83 ± 1.12

[a] indicates families that ranked the same for both pipelines. [b] indicates families in the top 10 most abundant that ranked differently for each pipeline.

Table 6
Richness, alpha-, beta- and gamma-diversity at family level.

Diversity measure	MG-RAST	QIIME2	P-value
S_{obs}	30	38.95	n/a
γ	103	81	n/a
Chao1	37.06	38.95	0.0275
H'adj	0.56 ± 0.05	0.60 ± 0.05	1.49E-64
D'adj	0.81 ± 0.06	0.84 ± 0.05	2.72E-16
β_A	73	42.05	n/a
β_M	3.43	2.08	n/a
C_{BC}	0.35 ± 0.1	0.31 ± 0.08	6.05E-21

Observed richness, S_{obs}, refers to the average richness of each sample; Gamma, γ, refers to richness of all 188 samples; Chao1 refers to the average estimated number of families; H'adj and D′ adj refer to the adjusted Shannon and the Simpson indices, respectively; β_A and β_M refer to the additive and multiplicative beta-diversity partitioning, respectively; C_{BC} refers to the Bray-Curtis index. n/a indicates non-applicable. *P*-values were obtained from two-sided paired t-tests assuming different variances.

Forty eight of the 97 genera detected by both MG-RAST and QIIME2 were reported with significantly different relative abundances (Table S1). Furthermore, for 67 of these 97 genera, the relative abundances obtained from the different pipelines were significantly correlated, including 36 genera with correlations equal to or higher than 0.8.

The genus-level compositions of the 114 samples collected at the end of the trial from the caecum, colon and faeces ($N = 38$ each collection site) were used in PLS-DA analyses to discriminate microbiota communities based on their collection sites, and compare results obtained using different pipelines. These comparisons were based on different data scenarios (described in detail in Table 1).

The explanatory variables included in PLS-DA models are evaluated based on their Variable Importance in Projection (VIP) values (Lima et al., 2019; Mao et al., 2016; Martínez-Álvaro et al., 2020; Roehe et al., 2016), which represent the significance of each variable to discriminate between sample types (caecum, colon and faeces). Using data scenarios A (all genera reported by each pipeline) and C (only genera identified by both pipelines), MG-RAST had higher R^2 value for discriminating between sample collection sites than QIIME2 (A: 77.19% and 74.56%; C: 73.68% and 68.42%, respectively). In contrast, for data scenarios B and D (same criteria as used for data scenarios A and C, respectively, but additional exclusion of genera with average relative abundance <1%), the genera obtained from MG-RAST resulted in lower R^2 than those obtained from QIIME2 (B: 61.40% and 72.81%; D: 58.77% and 62.28%, respectively; Table 7).

Although the number of genera resulting in VIP > 1 differed, their proportion in relation to the total number of genera was very similar (Table 8). Specifically, using data scenario A, the number of genera with VIP > 1 was higher when using genera compositions obtained from MG-RAST than those from QIIME2 (62 and 43, respectively). However, it corresponded to very similar proportions of genera considered in the models (28% and 27%). The analysis of data scenario B, revealed that 4 and 8 genera (29% and 36% of the genera considered in the model) resulted in VIP > 1 using MG-RAST and QIIME2, respectively. Corresponding results using scenario C were 24 and 22 genera (28% and 26%) when using genera compositions obtained from MG-RAST and QIIME2, respectively. In scenario D, both pipelines resulted in 3 genera (30%) with VIP > 1: *Ruminococcus*, *Blautia* and *Faecalibacterium,* whereas *Phascolarctobacterium* and *Lactobacillus* had the lowest VIP values in both pipelines (Table 9).

The most significant variables (VIP > 1) obtained from the PLS-DA analyses of the datasets obtained from MG-RAST and QIIME2 were not concordant (Fig. 6). These results suggest that the genera discriminating between sample types (caecum, colon and faeces) differ according to the pipeline applied for taxonomic annotation.

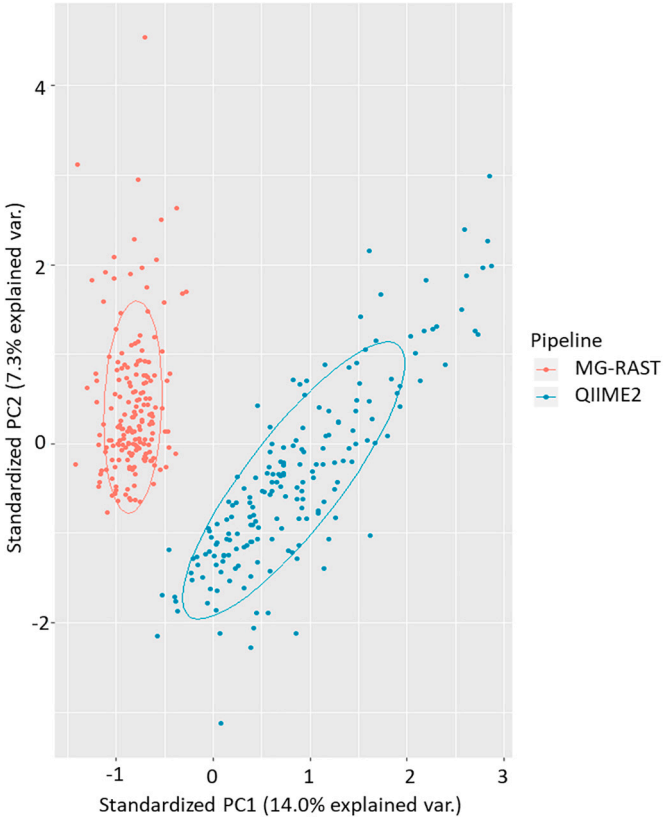


Fig. 4. Principal Components Analysis (PCA) of the relative abundances of the 56 families retrieved from both MG-RAST and QIIME2 (red and blue, respectively). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

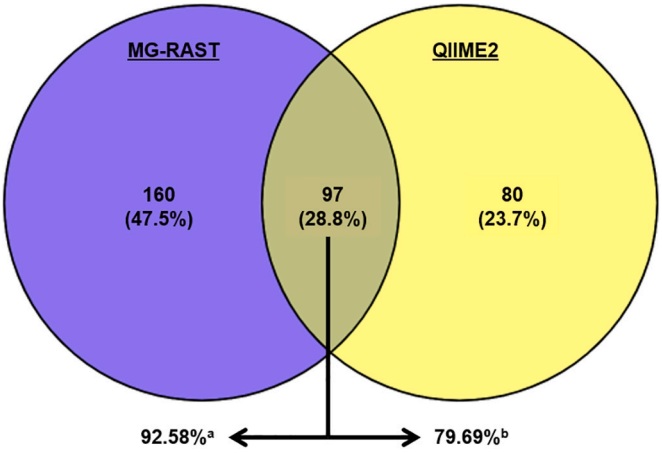


Fig. 5. Commonly and differently identified genera using MG-RAST and QIIME2. The numbers in brackets represent the proportion of in each area. Ninety seven genera were identified by both pipelines and they correspond to (a) 92.58% of the sequence hits reported by MG-RAST and (b) 79.69% of the sequence hits reported by QIIME2.

Co-abundance network analyses were carried out using the taxonomic compositions of the samples as described for the different filtering scenarios (Table 1, Fig. 7). For scenarios A, B and C we found significant enrichments of the clusters according to the pipeline from which the taxonomic compositions were generated (e.g. clusters 1 were consistently enriched in compositions from MG-RAST, whereas clusters 2 and 3 were consistently enriched in those obtained from QIIME2). However, using data scenario D, enrichments were observed to be related to the sample collection sites (e.g. cluster 1 was enriched in colon samples, clusters 3 and 5 were enriched in faeces samples and cluster 4 was enriched in caecum samples), whereas no significant enrichment of the pipelines was observed (i.e. samples collected from the same sites clustered together independently of the pipeline in which they were processed). These results concur with the PCA plots in Fig. 8, where, for data scenario C, the clear separation between samples is consistent with the pipeline employed, whereas for scenario D no distinct clusters were observed.

4. Discussion

Bioinformatics tools such as MG-RAST and QIIME2 are crucial for the characterization of environmental microbiota based on the 16S rRNA amplicon sequences retrieved from samples such as those from the swine gut content. We compared MG-RAST and QIIME2 because they are both self-contained pipelines that have been widely used for the identification of microbial communities from 16S rRNA amplicon sequences. Our research highlights differences in microbiota profiles retrieved from these tools on domain-, phylum-, family- and genus-level taxonomic compositions, while considering several data filtering procedures with focus on genus-level.

MG-RAST is an automated web-based tool that associates a priority level to the submitted project depending on its privacy parameters, which are defined by the user (priority level increases with decreased privacy settings). The duration of the analyses depends on the size of the submitted dataset, on the number of projects submitted to the MG-RAST

Table 8
Number and percentage of variables with variable importance in projection (VIP) greater than 1.

Data scenario	Total explanatory variables imputed		Explanatory variables with VIP > 1 (number and percentage)			
	MG-RAST	QIIME2	MG-RAST	(%)	QIIME2	(%)
A	225	159	62	27.56	43	27.04
B	14	22	4	28.57	8	36.36
C	86	86	24	27.91	22	25.58
D	10	10	3	30.00	3	30.00

Table 7
Confusion matrix for Partial Least Squares-Discriminant Analyses (PLS-DA) calculated using 114 samples collected from different body sites. PLS-DA were performed to discriminate between sample types, using 4 data scenarios, which included: (A) all genera reported by MG-RAST and QIIME2 (225 and 159, respectively); (B) the genera with relative abundance ≥1% reported by MG-RAST and QIIME2 (22 and 14, respectively); (C) the 86 genera identified by both pipelines and (D) the 10 genera with relative abundance ≥1% identified by both pipelines.

Data scenario	Predicted	Observed				Correctly assigned (%)	
		Caecum (N = 38)	Colon (*N* = 38)	Faeces (N = 38)	Subtotal	per collection site	per data scenario
MG-RAST							
	Caecum	20	4	0	24	52.63	
A	Colon	17	34	4	55	89.47	77.19
	Faeces	1	0	34	35	89.47	
	Caecum	27	18	3	48	71.05	
B	Colon	8	12	4	24	31.58	61.40
	Faeces	3	8	31	42	81.58	
	Caecum	22	3	0	25	57.89	
C	Colon	14	33	9	56	86.84	73.68
	Faeces	2	2	29	33	76.32	
	Caecum	24	17	2	43	63.16	
D	Colon	13	14	7	34	36.84	58.77
	Faeces	1	7	29	37	76.32	
QIIME2							
	Caecum	26	10	0	36	68.42	
A	Colon	11	24	3	38	63.16	74.56
	Faeces	1	4	35	40	92.11	
	Caecum	30	14	0	44	78.95	
B	Colon	6	21	6	33	55.26	72.81
	Faeces	2	3	32	37	84.21	
	Caecum	27	17	0	44	71.05	
C	Colon	10	19	6	35	50.00	68.42
	Faeces	1	2	32	35	84.21	
	Caecum	28	18	2	48	73.68	
D	Colon	8	11	4	23	28.95	62.28
	Faeces	2	9	32	43	84.21	

Table 9
Genus Variable Importance in Projection (VIP) calculated by Partial Least Squares – Discriminant Analysis including variables with average relative abundance superior to 1% identified by MG-RAST and QIIME2 (data scenario D).

MG-RAST	Average relative abundance (%)			VIP
Genus	Caecum	Colon	Faeces	
Ruminococcus[a]	3.22	3.76	5.78	1.74
Blautia	1.84	2.02	3.23	1.32
Faecalibacterium	17.47	19.20	13.10	1.25
Megasphaera	2.65	3.05	2.91	0.89
Clostridium	13.28	8.24	15.80	0.84
Prevotella	49.62	50.64	45.10	0.81
Dialister	3.82	3.95	2.95	0.75
Catenibacterium	1.76	2.08	2.33	0.66
Phascolarctobacterium[a]	1.91	2.06	2.44	0.61
Lactobacillus[a]	4.42	4.98	6.36	0.39

QIIME2	Average relative abundance (%)			VIP
Genus	Caecum	Colon	Faeces	
Ruminococcus[a]	2.31	2.33	4.31	1.79
Faecalibacterium	9.48	8.96	5.18	1.53
Blautia	2.84	3.29	4.08	1.16
Dialister	5.94	5.36	4.17	0.94
Catenibacterium	3.89	3.47	3.82	0.72
Megasphaera	3.35	3.24	3.43	0.71
Clostridium	6.11	6.90	8.97	0.61
Prevotella	57.00	56.98	53.59	0.60
Phascolarctobacterium[a]	2.74	2.27	3.11	0.52
Lactobacillus[a]	6.33	7.21	9.34	0.47

[a] indicates genus with the same VIP rank between the 2 pipelines.

server at the time of submission and on the server's characteristics. After the feature annotation, MG-RAST allows the user to alter data cleaning parameters maximum e-value, minimum percent similarity, minimum length of alignment and minimum abundance. This constitutes a major advantage for MG-RAST, as it provides the user the possibility of, in a manner of minutes, obtaining several versions of the calculated taxonomic tables. Due to the fact that the taxonomic annotations are saved in the server indefinitely, this process is feasible, fast and extremely practical, particularly because there is no gold standard when it comes to these parameters, so at any time during the analysis, the user is free to go back and test more appropriate parameters potentially generating more accurate results (Randle-Boggis et al., 2016).

QIIME2 is an installable set of dependencies that can be used for analysis as soon as the installation is finished. The analysis run time will depend on the size of the submitted dataset as in the case of MG-RAST, but also on the computer's characteristics and the user's bioinformatics skills. Unlike MG-RAST, QIIME2 does not provide inherent data hosting and sharing capabilities, placing the onus on the bioinformatician to disseminate the resulting (often large) files. QIIME2 does however provide web interfaces allowing users to examine QIIME2 output files without the need for installed software. Furthermore, in QIIME2, if the user desires to tune parameters such as minimum alignment length or e-value, an analysis will need to be performed from scratch, unless the intermediate files (which again can be large) are maintained by the user. To aid with rerunning microbiome analysis from intermediate files, output artefacts produced by QIIME2 contain a record of their prior processing steps (i.e. their 'provenance'). However, modifying microbiome analysis can still be cumbersome in comparison with the ease of MG-RAST.

4.1. Phylum- and family-level analyses

The percentage of unclassified and unidentified sequences on the phylum-level analyses was significantly higher in taxonomic compositions obtained from MG-RAST, in agreement with results of phylum-level analysis in Plummer et al. (2015), and family-level analysis in D'Argenio et al. (2014), although both these studies compared MG-RAST to QIIME, rather than to QIIME2. However, whereas D'Argenio et al. (2014) and Plummer et al. (2015) did not find statistically significant differences between family- and phylum-level relative abundances, respectively, 13 out of 15 phyla detected by both tools in this study had significantly different relative abundances. Additionally, at family-level, D'Argenio et al. (2014) did not find significantly different microbiota diversity (evaluated using Shannon index) resulting from MG-RAST and QIIME, whereas we observed QIIME2 to provide, on average, a significantly more even composition than MG-RAST.

Seven phyla were identified exclusively by QIIME2 but not by MG-RAST, and five of them were prevalent in more than half of the samples. Of this group, *Epsilonbacteraeota*, *Planctomycetes*, *Deferribacteres* and WPS-2 have previously been identified in microbiota sampled from the swine gastro-intestinal tract (Burrough et al., 2015; Gresse et al., 2019; Han et al., 2018; Tan et al., 2019); but to the best of our knowledge, *Patescibacteria*, *Kiritimatiellaeota* and SAR have not. The identification of SAR could be interpreted as a false positive, mostly because of its low overall prevalence of 2% (corresponding to 4 of 112 faecal samples), and low relative abundance of 0.01% (calculated within the 4 samples). *Patescibacteria* and *Kiritimatiellaeota* were detected in more than 80% of the samples, with relative abundances of 0.12% and 0.81%, respectively. Although the presence of these phyla was not confirmed by the MG-RAST results, their prevalence and relative abundances (calculated within the samples in which they were detected, based on QIIME2 results) are higher than several other phyla identified by both tools, such as *Chlamydiae*, *Fusobacteria*, *Verrucomicrobia*, *Elusimocrobia* and *Synergistetes*. The results regarding these two phyla are most likely the reflexion of the finer-scale resolution of reads into ASVs provided by DADA2 within QIIME2.

Using QIIME2 and MG-RAST approaches collectively identified 128 families, with 56 detected by both pipelines, corresponding to 99.95% and 89.67% of the total hits, respectively. Therefore, the families exclusively reported by either of the pipelines represent only a small percentage of the actual number of hits. In agreement with Kaszubinski et al. (2019), our results regarding alpha- and gamma-diversity at the family-level revealed that taxonomic compositions obtained from QIIME2 had on average higher observed richness (S_{obs}) but lower overall richness (γ) than MG-RAST's counterparts. These results agree with expectations, since MG-RAST's algorithm is based on the clustering of

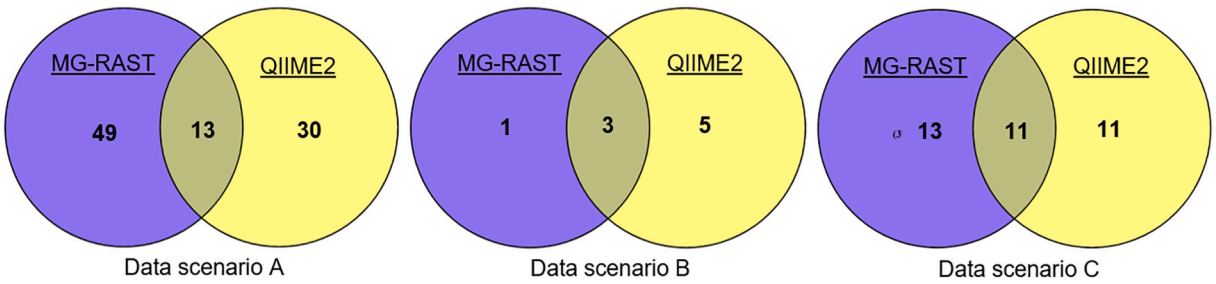


Fig. 6. Common and different genera with VIP > 1 as obtained from PLS-DA using data scenarios A, B and C.

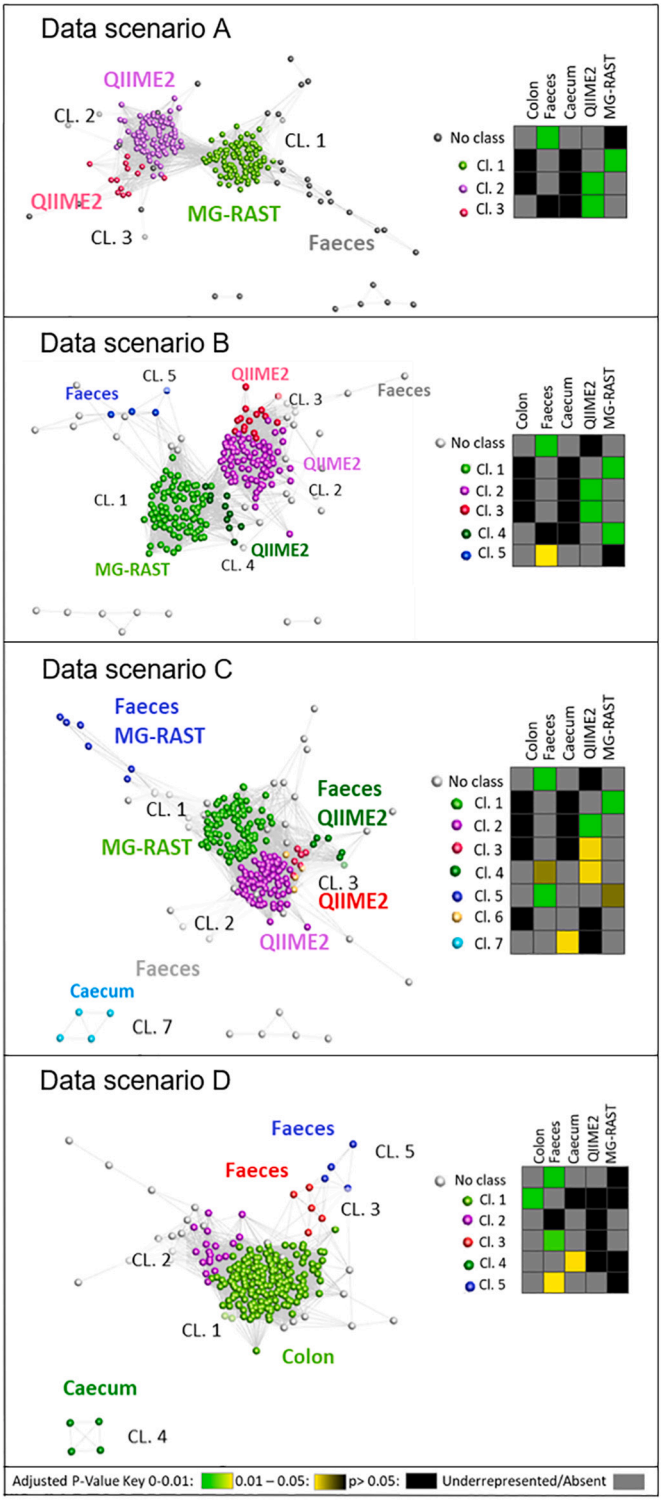


Fig. 7. Co-abundance networks and corresponding enrichment heat maps for samples based on data scenarios A (all taxa reported by each pipeline), B (all taxa reported by each pipeline with average relative abundance ≥1%), C (all taxa reported by both pipelines) and D (all taxa reported by both pipelines with average relative abundance ≥1%). Enrichment significance for each network is indicated by the heat maps (colour code can be found at the bottom of the figure). Nodes represent samples and edges correspond to correlations ≥0.95. CL. indicates cluster. Coloured names indicate clusters enrichment for the pipeline/sample collection sites.

sequences at 97% of similarity, which for a 254 bp length region such as the V4 region of the 16S rRNA gene would correspond to up to 8 Hamming distances, whereas DADA2, implemented in QIIME2, constructs an error model that is trained on the sequencing run and applied to create ASVs, resolving sequences with as low as 1 Hamming distance to the consensus sequence. As a consequence, it is expected that ASV-based algorithms better resolve fine-scale variation, leading to fewer spurious identifications than OTU-based methods (Callahan et al., 2016; Nearing et al., 2018; Prodan et al., 2020).

Beta-diversity measures suggested a higher similarity between the microbiota profiles computed in QIIME2 in comparison to MG-RAST, contrary to what was observed in Kaszubinski et al. (2019), possibly due to the filtering process applied, i.e. rarefaction to 1000 sequences to account for the variability of library size among pipelines, or to the removal of taxa with <1% relative abundance, leading to a loss of relevant information, or both (Kaszubinski et al., 2019).

4.2. Genus-level analyses

The analysis of the genus-level taxonomic compositions generated with MG-RAST and QIIME2 revealed that 97 (out of a total of 337) genera were identified by both pipelines and corresponded to 92.58% and 79.69% of the total number of sequence hits, respectively. Additionally, although the relative abundances of 67 genera identified by both tools were significantly correlated, only 36 genera whose abundances showed correlations of 0.8 or higher, suggesting that even though the majority of the read annotations concurred between the pipelines, the relative abundances of the communities were substantially different, which could have meaningful consequences in subsequent statistical analyses.

Furthermore, PLS-DA and co-abundance network analyses were used here in an attempt to discriminate sampling sites, and showed better accuracies for MG-RAST when using the whole taxonomic compositions (data scenarios A and C), whereas results were better for QIIME2 when using the minimum relative abundance threshold of 1% (data scenarios B and D), highlighting the importance of low abundance taxa, particularly in quantitative studies focused on associations between a host phenotype and its microbiota. It follows that if low abundance genera are real observations, MG-RAST is preferred to QIIME2, whereas if there is a large error involved in low abundance genera, QIIME2 would be the method of choice to determine the microbial community. In particular, results obtained from MG-RAST included identification of false positive taxa, and this is confirmed by microbiological knowledge, database information and the literature; however, inaccuracy in relative abundance of the identified taxa in the samples is more difficult to verify but this could be partly achieved by using mock community DNA samples.

The results obtained from scenarios B and D agreed with Kaszubinski et al. (2019), in which they found that random forest models applied to distinguish the sample area and death manner based on phylum- and family-level compositions had slightly higher mean prediction errors when using data obtained from MG-RAST than from QIIME2. The agreement between both studies may be due to the filtering procedures applied in Kaszubinski et al. (2019), i.e. rarefaction to 1000 sequences to account for the variability of library size among pipelines and removal of taxa with <1% relative abundance, which is similar to the restriction imposed on data scenarios B and D, where we filtered out all taxa with relative abundance lower than 1%. Additionally, the VIP (obtained from the PLS-DA analyses) associated to each genus differed substantially, revealing that the influence of the explanatory genera to the model differed, even when the same genera (but different relative abundances) were used, which could have implications in biomarkers identification. Additionally, due to the compositional nature of the microbiota datasets, we performed an exploratory analysis that compared the results of data scenarios A and B in PLS-DA models based on data transformed by the additive logratio methodology. The results were similar to those based on relative abundances, i.e. higher accuracy of discrimination between

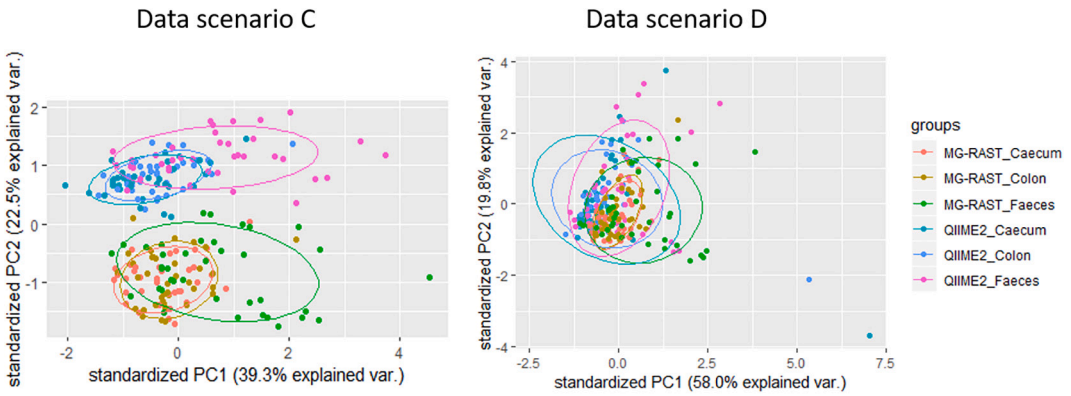


Fig. 8. Principal Components Analysis (PCA) plots for data scenarios C (86 genera taxa reported by both pipelines and corresponding relative abundances) and D (10 genera reported by both pipelines with a minimum average relative abundance of 1%).

sample types using MG-RAST than QIIME2 based on the whole datasets but substantially higher accuracy of QIIME2 when low abundance taxa (<1%) were removed.

5. Conclusions

This study found meaningful differences in the microbiota profiles generated by MG-RAST and QIIME2 from 16S rRNA amplicon sequences of samples from swine caecum, colon and faeces, not only in terms of which taxa were detected but also based on their relative abundances and overall prevalence. When low abundant (<1%) taxonomies are real and of importance, microbiota compositions generated by MG-RAST are preferred because it discriminated more accurately the collection sites of our samples than those generated by QIIME2. However, using the same accuracy criteria, QIIME2 is recommended when low abundance genera (<1%) were removed.

This research provides further evidence that the selection of pipeline greatly affects the outcomes and clarifies some of the characteristics of MG-RAST and QIIME2 which resulted in different taxonomic profiles based on the same amplicon sequence data. The user should be aware that, due to richness bias, employing an ASV-based pipeline may lead to increased false positives and thus artificially inflate alpha-diversity. However, the ASVs are a natural evolution for assessment of taxonomic profiles, whereas OTUs are based on an arbitrary similarity threshold, typically 97%. Additionally, irrespective of the pipeline selected, procedures such as trimming and parameters including minimum similarity to the taxonomic database should be tuned in each work, in order to maximize the pipelines' potential. The choice of pipeline and parameter tuning procedures should account for the context in which the resulting taxonomic compositions will be applied, whether this is for example, a study on human-derived post-mortem microbiome with forensic applications, in which fine-tuning of low abundance taxa is essential, or a study on cattle-derived rumen microbiome in association to animal production traits, in which high abundance taxa would potentially be more relevant.

Generally, we highlight the necessity to consider the potentially large differences in microbial community compositions based on different pipelines in the interpretation of results of microbiome studies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

www.metaplat.euThis work was supported by Zoetis Belgium SA, The Scottish Government, Biotechnology and Biological Sciences Research Council, United Kingdom (BBSRC BB/N01720X/1 and BB/S006567/1), and the MetaPlat project, European Union (www.metaplat.eu) H2020-MSCA-RISE-2015. The authors thank Professor Folker Meyer for helpful discussion.

References

Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26, 32–46. https://doi.org/10.1046/j.1442-9993.2001.01070.x.

Andrade, B.G.N., Bressani, F.A., Cuadrat, R.R.C., Tizioto, P.C., De Oliveira, P.S.N., Mourão, G.B., Coutinho, L.L., Reecy, J.M., Koltes, J.E., Walsh, P., Berndt, A., Palhares, J.C.P., Regitano, L.C.A., 2020. The structure of microbial populations in Nelore GIT reveals inter-dependency of methanogens in feces and rumen. J. Anim. Sci. Biotechnol. 11, 1–10. https://doi.org/10.1186/s40104-019-0422-x.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K. Bin, Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A.V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson II, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C. H., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., Caporaso, J. G., 2018. QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Prepr. https://doi.org/10.7287/peerj.preprints.27295.

Burrough, E.R., Arruda, B.L., Patience, J.F., Plummer, P.J., 2015. Alterations in the colonic microbiota of pigs associated with feeding distillers dried grains with solubles. PLoS One 10, e0141337. https://doi.org/10.1371/journal.pone.0141337.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869.

Chao, A., 1987. Estimating the population size for capture-recapture data with unequal catchability. Biometrics 43, 783–791. https://doi.org/10.4081/cp.2017.979.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., Chandra, S., McGarrell, D.M., Schmidt, T.M., Garrity, G.M., Tiedje, J.M., 2003. The ribosomal database project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Res. 31, 442–443. https://doi.org/10.1093/nar/gkg039.

Cox, M.P., Peterson, D.A., Biggs, P.J., 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11, 485. https://doi.org/10.1016/j.ophtha.2011.10.010.

D'Argenio, V., Casaburi, G., Precone, V., Salvatore, F., 2014. Comparative metagenomic analysis of human gut microbiome composition using two different bioinformatic pipelines. Biomed. Res. Int. 2014 https://doi.org/10.1155/2014/325340.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072. https://doi.org/10.1128/AEM.03006-05.

Development Team, 2020. Qiime2 Docs: Denoise-Paired: Denoise and Dereplicate Paired-End Sequences [WWW Document]. URL. https://docs.qiime2.org/2020.2/plugins/available/dada2/denoise-paired/.

Freeman, T.C., Goldovsky, L., Brosch, M., van Dongen, S., Maziere, P., Grocock, R.J., Freilich, S., Thornton, J., Enright, A.J., 2007. Construction, visualisation, and clustering of transcription networks from microarray expression data. PLoS Comput. Biol. 3, e206 https://doi.org/10.1371/journal.pcbi.0030206.

Glass, E.M., Wilkening, J., Wilke, A., Antonopoulos, D., Meyer, F., 2010. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harb Protoc 2010. https://doi.org/10.1101/pdb.prot5368.

Greenacre, M., 2018. Chapter 3: Logratio transformations. In: Compositional Data Analysis in Practice, pp. 17–24.

Gresse, R., Durand, F.C., Dunière, L., Blanquet-Diot, S., Forano, E., 2019. Microbiota composition and functional profiling throughout the gastrointestinal tract of commercial weaning piglets. Microorganisms 7, 343. https://doi.org/10.3390/microorganisms7090343.

Hadley, A., Hester, J., Chang, W., Hester, M.J., Wickham, H., Hester, J., Chang, W., 2019. Package 'devtools'.

Han, G.G., Lee, J.Y., Jin, G.D., Park, J., Choi, Y.H., Kang, S.K., Chae, B.J., Kim, E.B., Choi, Y.J., 2018. Tracing of the fecal microbiota of commercial pigs at five growth stages from birth to shipment. Sci. Rep. 8 https://doi.org/10.1038/s41598-018-24508-7.

Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. Genome Res. 17, 377–386. https://doi.org/10.1101/gr.5969107.

Kaszubinski, S.F., Pechal, J.L., Schmidt, C.J., Heather, R.J., Benbow, M.E., Meek, M.H., 2019. Evaluating bioinformatic pipeline performance for forensic microbiome analysis. J. Forensic Sci. https://doi.org/10.1111/1556-4029.14213.

Kent, W.J., 2002. BLAT - the BLAST-like alignment tool. Genome Res. 12, 656–664. https://doi.org/10.1101/gr.229202. Article published online before March 2002.

Koringa, P.G., Thakkar, J.R., Pandit, R.J., Hinsu, A.T., Parekh, M.J., Shah, R.K., Jakhesara, S.J., Joshi, C.G., 2019. Metagenomic characterisation of ruminal bacterial diversity in buffaloes from birth to adulthood using 16S rRNA gene amplicon sequencing. Funct. Integr. Genomics 19, 237–247. https://doi.org/10.1007/s10142-018-0640-x.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Le Cao, K.-A., Rohart, F., Gonzalez, I., Dejean, S., Abadi, A., Gautier, B., Bartolo, F., Monget, P., Coquery, J., Yao, F., Liquet, B., 2020. Package 'mixOmics'.

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

Li, F., Zhou, M., Ominski, K., Guan, L.L., 2016. Does the rumen microbiome play a role in feed efficiency of beef cattle? J. Anim. Sci. 94, 44–48. https://doi.org/10.2527/jas.2016-0524.

Lima, J., Auffret, M.D., Stewart, R.D., Dewhurst, R.J., Duthie, C.-A., Snelling, T.J., Walker, A.W., Freeman, T.C., Watson, M., Roehe, R., 2019. Identification of rumen microbial genes involved in pathways linked to appetite, growth, and feed conversion efficiency in cattle. Front. Genet. 10, 701. https://doi.org/10.3389/fgene.2019.00701.

Mao, S.-Y., Huo, W.-J., Zhu, W.-Y., 2016. Microbiome-metabolome analysis reveals unhealthy alterations in the composition and metabolism of ruminal microbiota with increasing dietary grain in a goat model. Environ. Microbiol. 18, 525–541. https://doi.org/10.1111/1462-2920.12724.

Martínez-Álvaro, M., Auffret, M.D., Stewart, R.D., Dewhurst, R.J., Duthie, C.A., Rooke, J. A., Wallace, R.J., Shih, B., Freeman, T.C., Watson, M., Roehe, R., 2020. Identification of complex rumen microbiome interaction within diverse functional niches as mechanisms affecting the variation of methane emissions in bovine. Front. Microbiol. 11, 659. https://doi.org/10.3389/fmicb.2020.00659.

Mcardle, B.H., Anderson, M.J., 2001. Fitting multivariate models to community data : a comment on distance-based redundancy analysis. Ecology 82, 290–297.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R., 2008. The Metagenomics RAST server: a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9 (386), 325–331. https://doi.org/10.1002/9781118010518.ch37.

Nearing, J.T., Douglas, G.M., Comeau, A.M., Langille, M.G.I., 2018. Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ 2018, 1–22. https://doi.org/10.7717/peerj.5364.

Oksanen, A.J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., Hara, R.B.O., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. Package 'vegan'.

Oliveros, J.C., 2007. Venny. An interactive tool for comparing lists with Venn's diagrams [WWW document]. URL. https://bioinfogp.cnb.csic.es/tools/venny/index.ht.

Plummer, E., Twin, J., Bulach, D.M., Garland, S.M., Tabrizi, S.N., 2015. A comparison of three bioinformatics pipelines for the analysis of preterm gut microbiota using 16S rRNA gene sequencing data. J. Proteomics Bioinform. 8, 283–291. https://doi.org/10.4172/jpb.1000381.

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A.H., Nieuwdorp, M., Levin, E., 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS One 15, e0227434. https://doi.org/10.1371/journal.pone.0227434.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glöckner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35, 7188–7196. https://doi.org/10.1093/nar/gkm864.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219.

Randle-Boggis, R.J., Helgason, T., Sapp, M., Ashton, P.D., 2016. Evaluating techniques for metagenome annotation using simulated sequence data. FEMS Microbiol. Ecol. 92 https://doi.org/10.1093/femsec/fiw095 fiw095.

Roehe, R., Dewhurst, R.J., Duthie, C.A., Rooke, J.A., McKain, N., Ross, D.W., Hyslop, J.J., Waterhouse, A., Freeman, T.C., Watson, M., Wallace, R.J., 2016. Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. PLoS Genet. 12, e1005846 https://doi.org/10.1371/journal.pgen.1005846.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584. https://doi.org/10.7717/peerj.2584.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09.

Tan, Z., Dong, W., Ding, Y., Ding, X., Zhang, Q., Jiang, L., 2019. Changes in cecal microbiota community of suckling piglets infected with porcine epidemic diarrhea virus. PLoS One 14, e0219868. https://doi.org/10.1371/journal.pone.0219868.

Vavrek, M.J., 2015. Package "fossil".

Veech, J.A., 2017. Measuring biodiversity. In: Encyclopedia of the Anthropocene. Elsevier Inc., pp. 287–295. https://doi.org/10.1016/b978-0-12-809665-9.10296-4

Vu, V.Q., 2011. Package "ggbiplot".

Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E.M., Kyrpides, N., Mavrommatis, K., Meyer, F., 2012. The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. BMC Bioinformatics 13. https://doi.org/10.1186/1471-2105-13-141.

Yu, Z., Morrison, M., 2004. Improved extraction of PCR-quality community DNA from digesta and fecal samples. Biotechniques 36, 808–812. https://doi.org/10.2144/04365st04.

2.3 Conclusions

The taxonomic characterization of samples based on the 16S rRNA gene carries many advantages in comparison to laboratory grown cultures. The wide range of microenvironments found in the gastrointestinal tract of animals makes it difficult for these conditions to be replicated using laboratory standard techniques, and so a large percentage of microorganisms are unculturable. For example, McCabe et al. (2015) suggested that less than 1% of the microorganisms in the rumen microbiota were culturable in the laboratory. Culture-independent approaches, such as those based on the 16S rRNA gene, allow for the characterization of the samples at low taxonomic levels (i.e., the genus level). Using specific primers for the amplification of the 16S rRNA gene in Prokaryotes, we derive the taxonomic composition based on the similarity of 16S rRNA amplicons to a reference database. 16S rRNA gene-based techniques are cost effective, have well-established protocols and bioinformatics pipelines, and well characterized reference databases (Ranjan et al., 2016). However, they also have limitations, for example, Ranjan et al. (2016) reported that this annotation technique is based on a putative association of the 16S rRNA gene with a taxon defined by an OTU. Additionally, the quantification of taxonomic groups based on the 16S rRNA gene is controversial, as the number of 16S rRNA gene copies in the genome of each group is variable, and therefore the abundance of 16S rRNA gene amplicon is not correlated with the abundance of the taxonomic groups (Smith and Osborn, 2009). Furthermore, bias can be introduced, for example, due to the variable affinity of the primers with their template and by the sequencing technologies used (Tremblay et al., 2015; Allali et al., 2017).

The advent of whole genome shotgun sequencing (WGS) allowed the development of techniques that overcome some of the disadvantages of the 16S rRNA gene-based approaches. These techniques are based on the sequencing of multiple random fragments of DNA within a sample, removing therefore the possible bias introduced by the use of primers. WGS has been shown to provide a more comprehensive characterization of the microbiome

profile present in a sample. Additionally, whereas the metagenomic profile can only be estimated when using 16S rRNA gene-based techniques, the WGS is able to identify specific genes, without the need for estimation (Ranjan et al., 2016).

Therefore, in our second manuscript we focused on the metagenomic composition derived from whole metagenomic sequencing of rumen microbiome samples, and on its association with production traits, including feed conversion efficiency, appetite, and growth rate in beef cattle.

2.4 References

Allali, I., Arnold, J. W., Roach, J., Cadenas, M. B., Butz, N., Hassan, H. M., et al. (2017). A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiol.* 17, 1–16. doi:10.1186/s12866-017-1101-8.

McCabe, M. S., Cormican, P., Keogh, K., O'Connor, A., O'Hara, E., Palladino, R. A., et al. (2015). Illumina MiSeq phylogenetic amplicon sequencing shows a large reduction of an uncharacterised succinivibrionaceae and an increase of the Methanobrevibacter gottschalkii clade in feed restricted cattle. *PLoS One* 10, 1–25. doi:10.1371/journal.pone.0133234.

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* 469, 967–977. doi:10.1016/j.bbrc.2015.12.083.

Smith, C. J., and Osborn, A. M. (2009). Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol. Ecol.* 67, 6–20. doi:10.1111/j.1574-6941.2008.00629.x.

Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., et al. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* 6, 1–15. doi:10.3389/fmicb.2015.00771.

Chapter 3 Identification of rumen microbial genes involved in pathways linked to appetite, growth, and feed conversion efficiency in cattle

3.1 Introduction

The second main objective of this thesis was to understand the association between the bovine rumen metagenome and host feed conversion efficiency traits, including feed conversion ratio (FCR), residual feed intake (RFI), average daily gain (ADG), and daily feed intake (DFI), in beef cattle; firstly, we aimed at evaluating the suitability of microbial genes in the rumen microbiome as biomarkers of host performance traits, and secondly, our goal was to identify which were the microbial genes more closely associated with these host traits, and in which biochemical pathways these microbial genes participated.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 August 2019
doi: 10.3389/fgene.2019.00701

Identification of Rumen Microbial Genes Involved in Pathways Linked to Appetite, Growth, and Feed Conversion Efficiency in Cattle

Joana Lima[1], Marc D. Auffret[1], Robert D. Stewart[2], Richard J. Dewhurst[1], Carol-Anne Duthie[1], Timothy J. Snelling[3], Alan W. Walker[3], Tom C. Freeman[2†], Mick Watson[2] and Rainer Roehe[1]**

[1] Beef and Sheep Research Centre, Future Farming Systems Group, Scotland's Rural College, Edinburgh, United Kingdom, [2] Division of Genetics and Genomics, The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, [3] The Rowett Institute, University of Aberdeen, Aberdeen, United Kingdom

OPEN ACCESS

Edited by:
Robert J. Schaefer,
University of Minnesota Twin Cities,
United States

Reviewed by:
Robert W. Li,
Agricultural Research Service
(USDA), United States
Prakash G. Koringa,
Anand Agricultural University,
India

***Correspondence:**
Joana Lima
Joana.Lima@sruc.ac.uk
Rainer Roehe
Rainer.Roehe@sruc.ac.uk

†ORCID:
Tom Freeman
orcid.org/0000-0001-5235-8483

Specialty section:
This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 07 February 2019
Accepted: 03 July 2019
Published: 08 August 2019

Citation:
Lima J, Auffret MD, Stewart RD,
Dewhurst RJ, Duthie C-A,
Snelling TJ, Walker AW, Freeman TC,
Watson M and Roehe R (2019)
Identification of Rumen Microbial
Genes Involved in Pathways Linked
to Appetite, Growth, and Feed
Conversion Efficiency in Cattle.
Front. Genet. 10:701.
doi: 10.3389/fgene.2019.00701

The rumen microbiome is essential for the biological processes involved in the conversion of feed into nutrients that can be utilized by the host animal. In the present research, the influence of the rumen microbiome on feed conversion efficiency, growth rate, and appetite of beef cattle was investigated using metagenomic data. Our aim was to explore the associations between microbial genes and functional pathways, to shed light on the influence of bacterial enzyme expression on host phenotypes. Two groups of cattle were selected on the basis of their high and low feed conversion ratio. Microbial DNA was extracted from rumen samples, and the relative abundances of microbial genes were determined *via* shotgun metagenomic sequencing. Using partial least squares analyses, we identified sets of 20, 14, 17, and 18 microbial genes whose relative abundances explained 63, 65, 66, and 73% of the variation of feed conversion efficiency, average daily weight gain, residual feed intake, and daily feed intake, respectively. The microbial genes associated with each of these traits were mostly different, but highly correlated traits such as feed conversion ratio and growth rate showed some overlapping genes. Consistent with this result, distinct clusters of a coabundance network were enriched with microbial genes identified to be related with feed conversion ratio and growth rate or daily feed intake and residual feed intake. Microbial genes encoding for proteins related to cell wall biosynthesis, hemicellulose, and cellulose degradation and host–microbiome crosstalk (e.g., *aguA, ptb*, K01188, and *murD*) were associated with feed conversion ratio and/or average daily gain. Genes related to vitamin B12 biosynthesis, environmental information processing, and bacterial mobility (e.g., *cobD*, *tolC*, and *fliN*) were associated with residual feed intake and/or daily feed intake. This research highlights the association of the microbiome with feed conversion processes, influencing growth rate and appetite, and it emphasizes the opportunity to use relative abundances of microbial genes in the prediction of these performance traits, with potential implementation in animal breeding programs and dietary interventions.

Keywords: feed conversion efficiency, appetite, metagenomics, rumen microbiome, microbial gene networks

INTRODUCTION

The global population is expected to reach 9.8 billion by 2050 (United Nations–Department of Economic and Social Affairs/Population Division, 2017), resulting in an escalation of the global demand for food and of the need for economically and environmentally sustainable livestock production systems (Godfray et al., 2010; Gerber et al., 2013). A large portion of livestock production is based on ruminants. In 2017, the EU-28 had a population of 88 million bovine animals, including cattle and water buffalo (Eurostat, 2018). Ruminants are particularly interesting due to their ability to convert human-indigestible plant biomass into high-quality products for human consumption such as meat and milk. Ruminants live in a symbiotic relationship with their rumen microbiota (comprising bacteria, protozoa, fungi, and archaea), which produce enzymes able to digest their food by breaking down complex polysaccharides of the plant biomass into volatile fatty acids (VFA), microbial proteins, and vitamins (Russell and Hespell, 1981; Bergman, 1990; Van Soest, 1994). Thus, the rumen microbiota fermentation profile has a significant influence on the feed conversion efficiency of the host (Russell, 2001; Li et al., 2009; Hernandez-Sanabria et al., 2011; Jami et al., 2014; Sasson et al., 2017; Meale et al., 2018) and is accountable for up to 70% of the host's daily energy requirements (Bergman, 1990).

In beef cattle production systems, expenses associated with feed account for up to 75% of the total production costs (Moran, 2005a; Nielsen et al., 2013), which makes the improvement of feed conversion efficiency very economically compelling. There is consequently great interest in understanding the host–microbial symbiotic relationships responsible for the conversion of feed into energy, protein, and vitamins usable by the host animal, but the mechanisms and degree to which the rumen microbiome impacts on animal production, health, and efficiency remain undercharacterized (Brulc et al., 2009; Creevey et al., 2014). Although the rumen harbors a core microbiome (Jami and Mizrahi, 2012; Henderson et al., 2015), in agreement with studies performed in the human gastrointestinal tract (Tap et al., 2009; Qin et al., 2010), the structure, and composition of the rumen microbiome varies within and between animals with differing performance traits. For example, in lactating dairy cattle, the increased methane yield during late lactation in comparison to early lactation within the same individual was found to be associated with significant changes in the ruminal microbial community structure (Lyons et al., 2018); Myer et al. (2015) showed different relative abundances of some microbial taxa and operational taxonomic units in animals with different average daily gain (ADG); Shabat et al. (2016) focused on residual feed intake (RFI) to demonstrate that highly efficient animals had a less diverse microbiota, being dominated by specific taxa and microbial genes which were involved in simpler metabolic pathway networks when compared to their less efficient counterparts. Other authors have reported that the rumen microbiome varies more between animals than within animals, proposing that the host itself and its physiological parameters have a significant influence on its own rumen microbiome (Li et al., 2009) and, therefore, on the efficiency of feed conversion into energy. In a mouse study, Benson et al. (2010) found that there is a well-defined portion of the gut microbiota that is subject to host genetic control, proposing it to be regarded as a host trait, rather than an environmental trait affecting the host. In agreement, in a beef cattle study, Roehe et al. (2016) confirmed the host genetic influence on the rumen bacterial composition using a genetic model based on sire progeny groups. The differences between sire progeny groups in methane emissions were in some cases larger than the differences found between diets differing largely in plant fiber content, suggesting a substantial host genetic influence on the microbial communities.

Selecting animals for breeding based on their ability to harvest energy from feed, together with nutritional interventions, could be the basis for an effective strategy to produce faster growing and more efficient animals (Gerber et al., 2013; Scollan et al., 2018). Given that the host has influence over the ruminal microbiome, which impacts the animals' feed conversion efficiency, this selection may be further improved by the inclusion of rumen metagenomic information into predictive models, as previously suggested by Ross et al. (2013). Feed conversion efficiency is very often estimated by either feed conversion ratio (FCR) or RFI; the latter is independent of growth and maturity patterns and is expected to be more sensitive and precise in measurements of feed utilization (Arthur and Herd, 2008). The use of microbial genes as proxies for feed conversion efficiency traits may be much more cost effective, rapid, and less labor intensive than their recording (Ross et al., 2013; Roehe et al., 2016). Our earlier research was the first proposing that the inclusion of relative abundance of microbial genes as proxies for FCR may be favorable, allowing their use as selection criteria for breeding animals, by identifying 49 microbial genes that explained 88.3% of the variation observed in FCR (Roehe et al., 2016). To our knowledge, no other studies have focused on the relationship between microbial gene abundances and RFI, daily feed intake (DFI), and ADG, which highlights the importance and novelty of the present work.

This study aimed at validating whether rumen microbial gene abundances are suitable proxies for feed conversion efficiency traits such as FCR; the analysis was further extended by focusing on RFI. Based on the previous evidence of strong interactions between the rumen microbiome and the host animal with consequences for feed conversion efficiency (Guan et al., 2008; Roehe et al., 2016; Shabat et al., 2016), we hypothesized that microbial gene abundances are linked to the animals' appetite and, consequently, to feed intake. A further aim of this research was to gain insight into the association of growth rate with the microbial gene abundances. Building on this, we aimed at better understanding the rumen microbial functional network associated with feed conversion efficiency and its component traits. This research will improve on the current knowledge about the impact of the rumen microbiome on appetite, growth, and efficiency of feed conversion processes.

MATERIALS AND METHODS

Ethics Statement

This study was conducted at the Beef and Sheep Research Centre, SRUC, UK. The study was carried out in accordance with the requirements of the UK Animals (Scientific Procedures) Act 1986. The protocol was approved by the Animal Experiment Committee of SRUC. All standard biosecurity and institutional

safety procedures were applied during the animal experiment and the laboratory analysis.

Animals, Adaptation Period, and Measurement of Traits

Two experiments were carried out to determine the effect of nitrate or lipid additives within different basal diets on methane emissions from beef cattle. The first experiment was conducted in 2013, and it consisted of a 2 × 2 × 3 factorial design including 84 steers of two breed types (crossbreed Charolais, CHx and Luing); two basal diets, forage (FOR) and concentrate (CONC), which consisted respectively of ratios of 520:480 and 84:916 forage to concentrate (g/kg dry matter); and three treatments, nitrate and lipid feed additives, as well as the control. From these animals, 24 animals were selected with extreme high and low FCR values within breed type and basal diet (two animals per feed additive and control). More details related to this experiment can be found in Duthie et al. (2015) and Troy et al. (2015). The second experiment was a 2 × 4 factorial design experiment, conducted in 2014, involving 80 animals. There were two breed types—40 crossbred Limousin (LIMx) and 40 crossbred Aberdeen Angus (AAx)—which were subject to a balanced design consisting of four dietary treatments using one basal diet (550:450 forage to concentrate ratio g/kg dry matter, FOR) and testing the effects of feed additives nitrate, lipid, or their combination in comparison to the control on methane output. Full details of the experiment are presented in Duthie et al. (2017). From this experiment, 18 animals were selected within each combination of breed type and diet: nine for the high FCR group and nine for the low FCR group. DFI was assessed by measuring dry matter intake (DMI, kg/day), which was recorded in both experiments using electronic feeding equipment (HOKO, Insentec, Marknesse, The Netherlands). Body weight (BW) was measured weekly using a calibrated weight scale (before fresh feed was offered). Growth was modeled by linear regression of BW against test date to obtain ADG, mid-test BW, and mid-test metabolic BW (MBW = $BW^{0.75}$). FCR was calculated as average DMI (kg/day) divided by ADG. RFI was estimated as deviation of actual DMI (kg/day) from DMI predicted based on linear regression of actual DMI on ADG, mid-MBW, and fat depth at $12^{th}/13^{th}$ rib at the end of the 56-day test (Duthie et al., 2015; Troy et al., 2015; Duthie et al., 2017).

A flowchart summarizing the methods for generation of data and subsequent statistical analyses is presented in **Figure 1**.

Sampling of Rumen Digesta and Whole Metagenomic Sequencing

As described in Duthie et al. (2015) and Auffret et al. (2017), animals from both experiments were slaughtered in a commercial

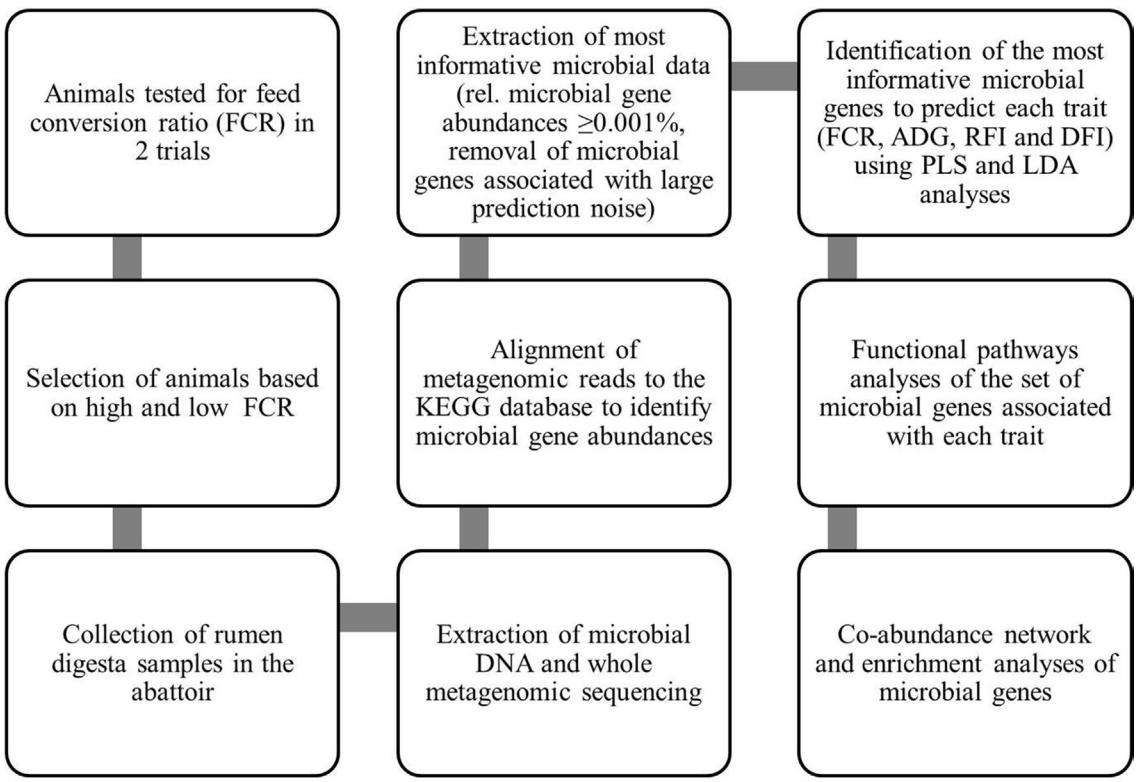


FIGURE 1 | Flowchart summarizing methods for generation of data and their statistical analyses: This flowchart summarizes how the data were generated and which statistical analyses were used to identify the associations between gene abundances and performance traits of animals to understand the rumen microbial functional pathways associated with these traits. KEGG, Kyoto Encyclopedia of Genes and Genomes; FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake; PLS, partial least squares; LDA, linear discriminant analysis.

abattoir where two samples of rumen digesta (~50 ml) were collected immediately after the rumen was opened to be drained. The slaughter house sample collection process results in well-mixed samples of rumen contents. DNA was extracted from the rumen samples of 42 animals following the methodology described in Rooke et al. (2014). Illumina TruSeq libraries were prepared from genomic DNA and sequenced on Illumina HiSeq systems 4000 by Edinburgh Genomics (Edinburgh, UK). Paired-end reads (2 × 150 bp) were generated, resulting in between 8 and 15 GB per sample (between 40 and 73 million paired reads). The raw data can be downloaded from the European Nucleotide Archive under accession PRJEB21624.

Identification of the Rumen Microbial Gene Abundances

Bioinformatics analysis for identification of rumen microbial genes was carried out as previously described by Wallace et al. (2015). Briefly, to measure the abundance of known functional microbial genes in the rumen samples, reads from whole metagenome sequencing were aligned to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000) using Novoalign (www.novocraft.com). Parameters were adjusted such that all hits were reported that were equal in quality to the best hit for each read and allowing up to a 10% mismatch across the fragment. The KEGG Orthologue groups (KO) of all hits that were equal to the best hit were examined. If we were unable to resolve the read to a single KO, the read was ignored; otherwise, the read was assigned to the unique KO. Read counts were summed and normalized to the total number of hits. This mapping of the whole metagenomic data to the KEGG database resulted in a dataset comprising of 4,966 KEGG genes. Microbial genes were removed from the dataset when they were absent from three or more animals and when the mean relative abundance was lower than 0.001%, leaving 1,692 microbial genes for further analyses.

Statistical Analysis

For each of the 1,692 microbial genes, a linear model was fitted, including as fixed effects a combined class variable of breed, diet, and year of experiment (six levels) and the FCR groups (high FCR, FCR-H and low FCR, FCR-L) using the lm() function in R version 3.4.2. The microbial genes which resulted in $P \geq 0.1$ for the differences in FCR groups were not considered in the partial least squares analyses (PLS, SAS version 9.3 for Windows, SAS Institute Inc., Cary, NC, USA) to avoid excessive noise of microbial genes uncorrelated to the traits of interest. In the linear model, FCR groups were replaced successively by ADG, RFI, and DFI as covariables to identify only potentially relevant microbial genes of these traits for further PLS analyses. In addition, genes with unknown function were removed from these datasets.

Microbial genes whose relative abundances were significantly associated to each trait in the linear models were analyzed using a sequential PLS-based methodology. First, PLS models were calculated in which the number of latent variables was determined by "leave-one-out" cross-validation, and genes with lower variable importance in projection (VIP) were removed. Second, the sets of genes created in the first step were evaluated by PLS models using three latent variables to determine the smaller set of genes leading to higher explained variation of both independent and dependent variables.

Each set of microbial genes identified in the PLS analyses as best predicting the trait was then used in a linear discriminant analysis (LDA), performed in R version 3.5.1 (2018-07-02) package MASS_7.3-51.4. In these analyses, the categories were for FCR those described previously as FCR-H and FCR-L; for all other traits, animals were classified as high or low, depending on their observations being higher or lower than the median (balanced for trial, breed, and diet).

The microbial genes identified to be significantly associated with each trait were submitted to an extensive review about their functionality based on databases such as KEGG (Kanehisa Laboratories, 2018), BioCyc (Karp et al., 2017), and UniProt (Bateman, 2019) and information from the literature.

Networks

The coabundances between microbial genes were investigated in a stepwise network analysis using the Graphia Professional software (Kajeka Ltd, Edinburgh; Freeman et al., 2007), in which nodes represent microbial genes and edges represent a correlation value above a defined value of *r*. In the first step, the correlation threshold of $r = 0.45$ was selected such that all microbial genes ($n = 1{,}692$) were included in the network. The microbial genes identified by PLS to be associated with a trait of interest were then located in the network. Clustering was performed using the Markov clustering method (MCL) available in Graphia Professional using the default settings (inflation, preinflation, and scheme values of 6). All clusters that held at least one microbial gene previously identified in the PLS analysis to be associated with a trait of interest were identified. These were incorporated into a new network generated at correlation threshold of $r = 0.80$ containing 1,135 microbial genes. MCL was then performed on this network, with inflation and preinflation values of 2 and scheme value of 6, reflecting the clustering structure suggested in the network itself. Analyses of enrichment of genes identified in the PLS as associated to each trait were performed on the clusters, and significance was assessed at $P < 0.05$.

RESULTS

Performance Traits Related With Feed Conversion Efficiency

The average FCR values observed for animals selected into FCR-H (inefficient) and FCR-L (efficient) groups differed significantly by 2.3 kg DFI/kg ADG (**Figure 2**). When comparing these two groups for other traits, the FCR-H group had significantly higher values of RFI (0.8 kg) and significantly lower ADG (0.39 kg); in the case of DFI, no significant difference was observed between the FCR groups.

ADG and FCR had a strong significant negative correlation of 0.80, suggesting that high growth rate is associated with efficient animals, using less feed per kilogram of weight gain. FCR and RFI were significantly positively correlated, but at a low level of

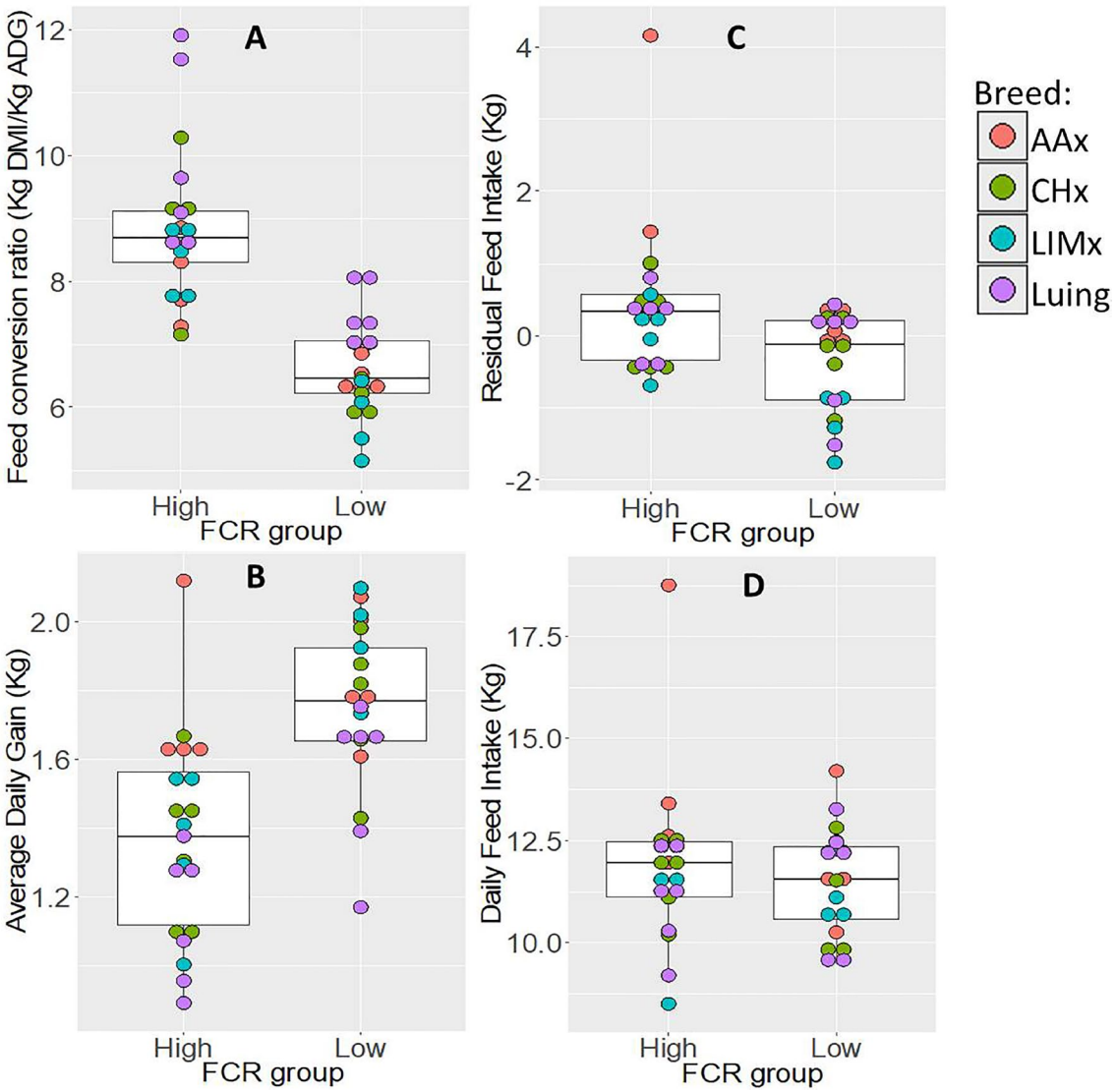


FIGURE 2 | Distribution of variation and range of performance traits: **(A)** feed conversion ratio, **(B)** average daily gain, **(C)** residual feed intake, and **(D)** daily feed intake within feed conversion ratio groups (high and low). The boxplots show the variation and range of each trait within each feed conversion ratio group. FCR, feed conversion ratio; AAx, crossbred Aberdeen Angus; CHx, crossbred Charolais; LIMx, crossbred Limousin.

0.32. DFI was significantly correlated with RFI and ADG at high and moderate levels of 0.77 and 0.53, respectively.

Rumen Microbial Genes Associated With Feed Conversion Efficiency Traits

The PLS analyses identified sets of 20 and 14 microbial genes whose relative abundances explained 63.4 and 65.4% of the variation in FCR and ADG, respectively, and sets of 17 and 18 microbial genes whose relative abundances explained 65.6 and 72.9% of the variation in RFI and DFI, respectively, including the combined fixed effect of diet, breed, and year of experiment (**Table 1**). Without this combined fixed effect, the variances explained by microbial genes in FCR and ADG decreased to 54.2 and 61.4%, while in RFI and DFI, they decreased to 50.8 and 67.7%, respectively. A discriminant analysis between groups of high- and low-performing animals, using the set of microbial genes identified in the PLS analysis to best predict each trait, resulted in prediction accuracies of 90, 79, 86, and 86% for FCR, ADG, RFI, and DFI (**Figure 3**).

The Venn diagram presented in **Figure 4** illustrates the overlap between the sets of genes identified for the prediction of each of the four traits. For the prediction of FCR and ADG, six microbial genes were simultaneously selected: UDP-N-acetylmuramoylalanine-D-glutamate ligase, glycine cleavage system H protein, translation initiation factor IF-1, N utilization substance protein A, DNA-binding protein HU-beta, and diphthamide synthase subunit *dph2* (*murD*, *gcvH*, *infA*, *nusA*, *hupB*, and *dph2*, respectively). Three microbial genes were

TABLE 1 | Percentage of variation in each trait explained by the microbial genes identified in the partial least squares (PLS).

		Percent variation accounted for by partial least squares factors			
		Model effects		Dependent variables	
Trait	No. factors	Current	Total	Current	Total
FCR	1	41.59	41.59	35.46	35.46
	2	6.35	47.94	21.19	56.65
	3	7.57	55.51	6.72	63.37
ADG	1	39.42	39.42	49.26	49.26
	2	9.60	49.02	11.47	60.73
	3	7.97	56.99	4.67	65.40
RFI	1	24.04	24.04	44.32	44.32
	2	13.95	37.99	16.80	61.12
	3	16.72	54.71	4.52	65.63
DFI	1	28.98	28.98	44.94	44.94
	2	21.25	50.23	19.94	64.88
	3	7.86	58.09	8.05	72.93

The number of factors refers to the number of latent variables in which the total number of microbial genes (independent variables) were projected in the PLS procedure, and each factor accounts for a portion of the total explained variation. The "Model Effects" columns refer to the percent variability of the independent variables matrix that relates to the respective percent variability presented in the "Dependent Variables" columns. The "Current" columns present values for each extracted factor individually, and the "Total" columns present the subtotal variation. The cells colored in gray contain the values of percent variation explained by the three latent variables for each trait. FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake.

simultaneously selected for the prediction of traits RFI and DFI: glucose-1-phosphate cytidylyltransferase, CDP-glucose 4,6-dehydratase, and energy-converting hydrogenase B subunit D (*rfbF*, *rfbG*, and *ehbD*, respectively). The microbial genes identified for the prediction of more than one trait are highlighted in the shaded rows in **Tables 2–5**, in which a more detailed information about their function and importance for prediction is provided.

Based on the relative abundance of 1,135 microbial genes across rumen samples, a coabundance network was developed (**Figure 5**), and clusters were identified. The clustering pattern evidences the microbial genes that are more closely connected to microbial genes previously identified in the PLS analyses. The network cluster to which each microbial gene belongs to is presented in **Tables 2–5**. Cluster 2 was significantly enriched for microbial genes predicting DFI and RFI&DFI (RFI and/or DFI). Cluster 4 was enriched for microbial genes predicting RFI and RFI&DFI. Microbial genes simultaneously predicting FCR and ADG were enriched in clusters 20 and 21, while those predicting FCR&ADG (FCR and/or ADG) were enriched in clusters 21 and 25. ADG-predicting microbial genes were enriched in clusters 21 and 25, whereas FCR-predicting genes were only enriched in cluster 25. Other genes previously identified in the PLS analysis were scattered across the graph.

Most microbial genes identified exclusively for the prediction of FCR are related to carbohydrate metabolism and transport: fructuronate reductase, galactokinase, alpha-glucuronidase, beta-glucuronidase, beta-glucosidase, phosphate butyryltransferase P, UDP-N-acetylglucosamine acyltransferase, gluconate 5-dehydrogenase, and lactate permease (respectively *uxuB*, *galK*, *aguA*, *uidA*, K01188, *ptb*, *lpxA*, *idnO*, and *lctP*) were proportionally more abundant in efficient animals (lower FCR, **Supplementary Figure S1A**). The microbial gene lactoylglutathione lyase (*glo1*) is also associated with carbohydrate metabolism and identified for predicting FCR, but it had higher relative abundance in less efficient animals (higher FCR). Microbial genes *galK* and *xylE* (i.e., MFS transporter, SP family, xylose:H+ symporter) were both located in cluster 5, but this cluster was not significantly enriched for microbial genes associated to FCR. On the other hand, cluster 25 was enriched due to the presence of microbial genes *uxuB* and *lpxA*.

Microbial genes associated with amino acid metabolism and transport pathways were identified for the prediction of ADG and found to be relatively more abundant in animals with higher ADG (see **Supplementary Figure S1B**), e.g., aspartate-semialdehyde dehydrogenase and phenylacetate-CoA ligase (*asd* and *paak*, respectively). Some housekeeping genes were also identified for this set, including large subunit ribosomal protein L17 and L36, F-type H+-transporting ATPase subunit delta and FKBP-type peptidyl-prolyl cis-trans isomerase *slyD* (*rplQ* and *rpmJ*, *atpH*, and *slyD*). Genes *rplQ*, *atpH*, and *slyD* were relatively more abundant in animals with higher ADG, and *rpmJ* was relatively more abundant in animals with lower ADG. The microbial gene N-acetylmuramoyl-L-alanine amidase (*amiABC*) was identified for prediction of ADG, being negatively correlated with the trait.

All microbial genes simultaneously identified for predicting FCR and ADG showed a negative correlation to FCR and a positive correlation to ADG. These included housekeeping genes (*infA*, *hupB*, and *dph2*), a gene related to carbohydrate metabolism (*gcvH*), *murD*, which was associated with peptidoglycan metabolism and D-glutamine and D-glutamate metabolism, and *nusA*, associated with transcription regulation. Cluster 21 was enriched in ADG- and FCR&ADG-predicting microbial genes due to the presence of *atpH*, *rplQ* (ADG), and *infA* (FCR&ADG).

Five microbial genes identified for the prediction of RFI were associated with environmental sensing, bacterial chemotaxis, and motility: sensor kinase *cheA*, response regulator *cheY*, methyl accepting chemotaxis protein, flagellar motor switch protein *fliN/fliY*, and flagellar hook protein *flgE* (*cheA*, *cheY*, *mcp*, *fliN*, and *flgE*, respectively) were found to be relatively more abundant in more efficient animals, i.e., lower RFI. Other microbial genes associated with RFI are involved in the biosynthesis of cofactors and vitamins, particularly vitamin B12 production, for example, cobalt transport protein, threonine-phosphate decarboxylase, and precorrin-6Y C5,15-methyltransferase (decarboxylating), which correspond respectively to *cbiN*, *cobD*, and *cobL* (**Supplementary Figure S1C**). Finally, three genes that encode proteins related to carbohydrate transport and metabolism were relatively more abundant in more efficient animals (i.e., lower RFI): the simple sugar transport system permease protein, oxaloacetate decarboxylase, alpha subunit, and aldehyde:ferredoxin oxidoreductase (respectively ABC.

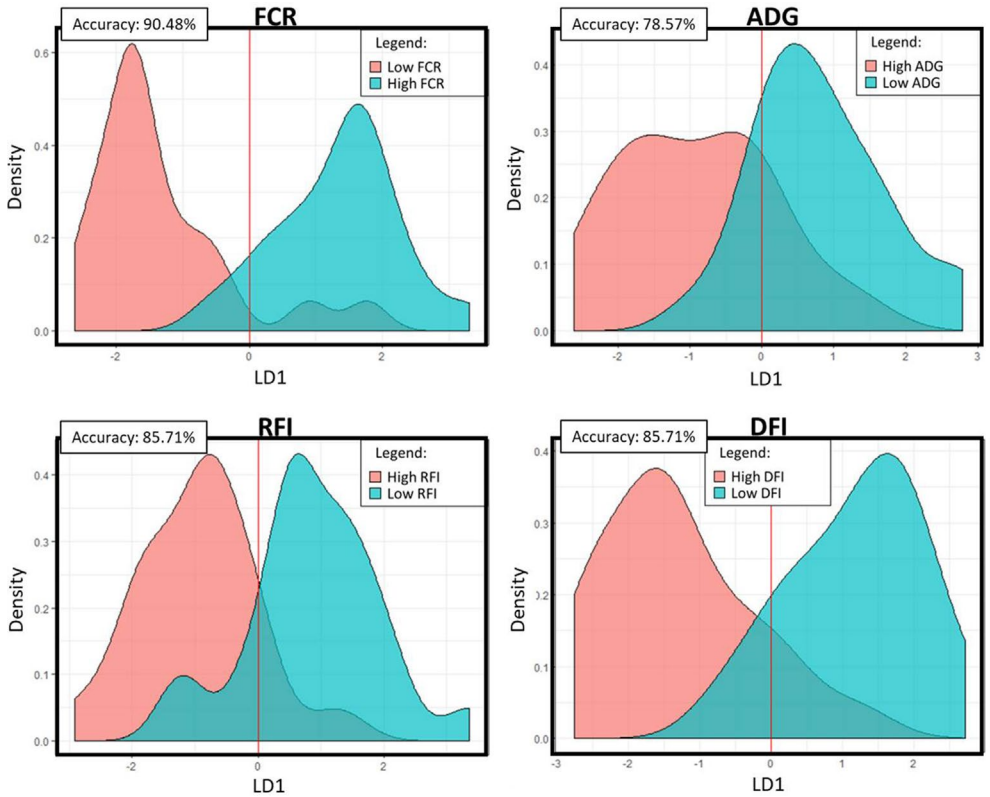


FIGURE 3 | Linear discriminant analysis density plots: Microbial genes identified in the PLS analyses to be significantly associated with the trait were used in a linear discriminant analysis of high- and low-performing animals. The density plots represent the predicted categories for each trait. The accuracy value represents the percentage of animals that were correctly assigned to their category. FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake; LD1, linear discriminant 1.

SS.P, *oadA*, and *aor*). Cluster 4 was significantly enriched in microbial genes associated with RFI due to the presence of microbial genes *cobD*, *cobL*, *mcp*, and *oadA*, and serine-type D-Ala-D-Ala carboxypeptidase (penicillin-binding protein 5/6), inner membrane protein, and Cd2+/Zn2+-exporting ATPase (respectively, *dacC*, *ybrG*, and *zntA*).

The set of microbial genes identified for prediction of DFI included four microbial genes, proportionally more abundant in animals with higher DFI, which encoded proteins associated with environmental sensing, i.e., nitrogen regulatory protein P-II 1, outer membrane channel protein *TolC*, and preprotein translocase subunit *YajC* (*glnB*, *tolC*, and *yajC*, respectively). Nitrate reductase 1, alpha subunit (*narG*) was related to denitrification, releasing nitrite, and it was found to be relatively more abundant in animals with lower DFI (**Supplementary Figure S1D**). DNA-directed RNA polymerase subunit beta (*rpoB*, proportional higher abundance in animals with lower DFI), ribosomal large subunit pseudouridine synthase B, exodeoxyribonuclease VII small subunit, ribonuclease III, N utilization substance protein B, and integration host factor subunit alpha (respectively *rluB*, *xseB*, *rnc*, *nusB*, and *ihfA*, proportionally more abundant in animals with higher DFI) are housekeeping genes identified in this work for the prediction of DFI. Cluster 2 was significantly enriched with microbial genes associated with DFI due to the presence of *glnB*, *infA*, *mrdA*, *nusB*, *rdgB*, *rluB*, *tolC*, and *xseB*.

RFI- and DFI-predicting genes include glucose-1-phosphate cytidylyltransferase, CDP-glucose 4,6-dehydratase (respectively *rfbF* and *rfbG*, related to amino sugar and nucleotide sugar metabolism), and energy-converting hydrogenase B subunit D (*ehbD*, housekeeping). These three genes were proportionally more abundant in less efficient animals (higher RFI associated with increased DFI).

DISCUSSION

Rumen Microbial Gene Abundances Associated With Efficiency Traits

Our research indicates that there is a substantial link between rumen microbial gene abundances and appetite (measured as feed intake), growth rate, and feed conversion efficiency (**Figure 6**). The relative abundances of 20 and 17 microbial genes accounted for substantial variation (>60%) in FCR and RFI, respectively. The discriminant analyses of high- and low-performing animals indicated that accurate classification (>85% correct assignment of FCR and RFI categories) could be achieved using the microbial genes identified in the PLS for the prediction of the traits. Roehe et al. (2016) also found an association of microbial gene abundances with FCR, but their results were based on a smaller number of animals selected for their extreme values in

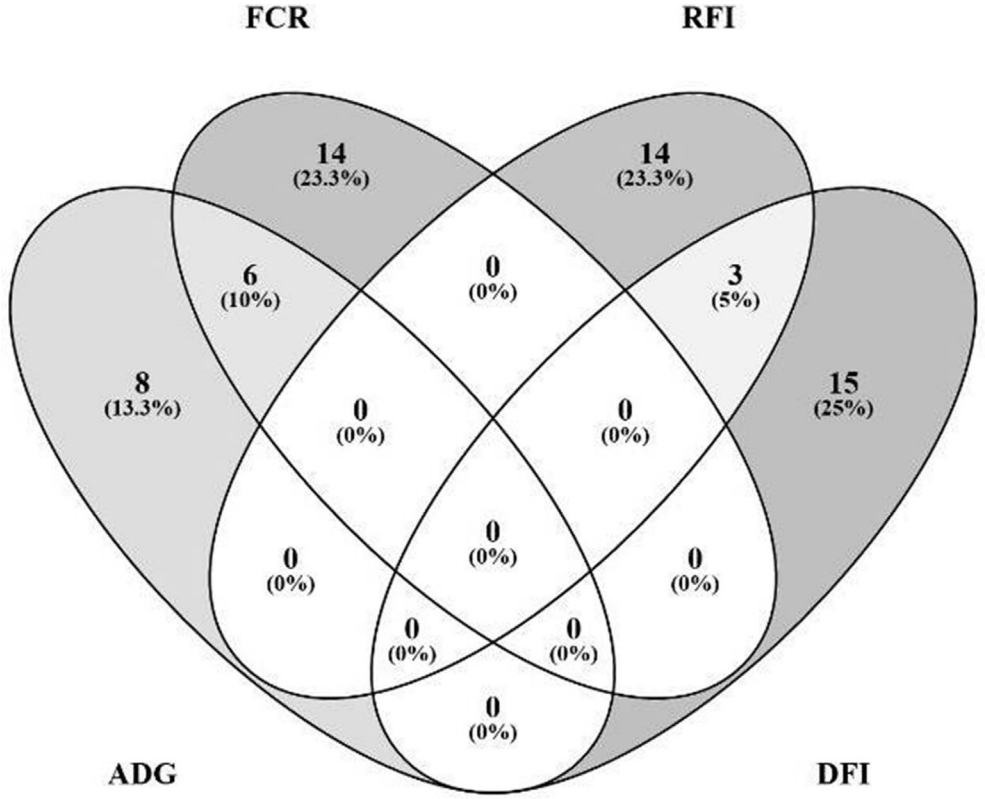


FIGURE 4 | Overlap analysis of identified microbial genes: The image illustrates the number of microbial genes identified in the partial least squares analysis as fitted for prediction of each animal performance trait exclusively, and the number of microbial genes simultaneously predicted for multiple traits: six microbial genes were simultaneously identified for prediction of FCR (feed conversion ratio) and ADG (average daily gain), and three for both RFI (residual feed intake) and DFI (daily feed intake).

methane emissions. In the present study, animals were selected based on their extreme FCR values, yielding a statistically more powerful estimate of this trait. Whereas FCR is calculated as a ratio between DFI and ADG and is therefore highly affected by growth rate and body composition, RFI is independent of these traits (Berry and Crowley, 2013). The low phenotypic correlation ($r = 0.32$) between FCR and RFI suggests that these traits capture substantially distinct characteristics.

For ADG and DFI, the relative abundances of 14 and 18 microbial genes, respectively, also explained substantial variation (>65%), and the discriminant analyses of high- and low-performing animals resulted in high prediction accuracies of 79 and 86%, respectively. These component traits were moderately correlated, agreeing with the report by Berry and Crowley (2013) of a large independent variation of feed intake and weight gain.

The animals' appetite, feeding behaviour, and gastrointestinal motility (among other traits) are thought to be regulated by several mechanisms, including a communication between the rumen microbiome and the brain, through the gut–liver–brain axis (vagus nerve). This communication has been proposed to be mediated by multiple mechanisms, such as insulin/glucagon homeostasis, oxidation of acetyl coenzyme A, and release of VFA by the rumen microbiota (like propionate, associated with hypophagic behavior in ruminants, or butyrate and acetate, associated with motility of the gastrointestinal tract in monogastric animals; Sakata and Tamate, 1979; Cherbut, 2003; Oba and Allen, 2003; Arora et al., 2011; Maldini and Allen, 2018). Given the predictability of performance traits using relative abundances of rumen microbial genes observed in the present research (particularly that of DFI) and the high impact of the rumen microbiome on feed intake regulation (as discussed in the literature), we hypothesize that rumen microbial genes are closely involved in the metabolic pathways that regulate feed intake.

Differential Microbial Gene Sets Predicting Distinct Trait Complexes

The coabundance microbial gene network (**Figure 5**) identified two separate trait complexes. While microbial genes identified for the prediction of FCR were grouped with ADG-predicting genes, microbial genes identified for the prediction of RFI were grouped with DFI-predicting genes, as revealed by differential enrichment in separate clusters (**Supplementary Figure S2**). For example, beta-glucosidase is encoded by microbial genes *bglX* and K01188, which were associated to different traits (DFI and FCR, respectively). This type of differential clustering was previously observed for microbial genes associated with methane

TABLE 2 | Summary of microbial genes identified for the prediction of FCR.

KEGG id	Description	Gene name abbreviation	Pathways	Mean abundance	PLS estimate	VIP	Cluster
K03783	Purine-nucleoside phosphorylase	*punA*	[2]Metabolic pathways; biosynthesis of secondary metabolites; purine metabolism; pyrimidine metabolism; nicotinate and nicotinamide metabolism	0.0107	–0.2755	1.22	[1]
K08138	MFS transporter, SP family, xylose:H+ symporter	*xylE*	[3]Carbohydrate transport and metabolism, amino acid transport and metabolism, Inorganic ion transport and metabolism	0.0404	0.1135	1.09	05
K00046	Gluconate 5-dehydrogenase	*idnO*	[4]L-idonate degradation	0.0845	0.0166	1.08	11
K00040	Fructuronate reductase	*uxuB*	[2]Metabolic pathways; pentose and glucuronate interconversions	0.0847	0.0503	1.01	25
K01759	Lactoylglutathione lyase	*glo1*	[2]Pyruvate metabolism	0.0021	0.1547	1.01	09
K00849	Galactokinase	*galK*	[2]Metabolic pathways; galactose metabolism; amino sugar and nucleotide sugar metabolism	0.0631	0.0675	1.00	05
K01195	Beta-glucuronidase	*uidA*	[2]Metabolic pathways; biosynthesis of secondary metabolites; pentose and glucuronate interconversions; glycosaminoglycan degradation; porphyrin and chlorophyll metabolism; flavone and flavonol biosynthesis; drug metabolism—other enzymes; lysosome	0.0127	–0.1174	0.99	07
K14220	tRNA Asn	*tRNA-Asn*	[2]Aminoacyl-tRNA biosynthesis	0.0155	0.0139	0.96	NC
K00677	UDP-N-acetylglucosamine acyltransferase	*lpxA*	[2]Metabolic pathways; lipopolysaccharide biosynthesis; cationic antimicrobial peptide (CAMP) resistance	0.0403	–0.0186	0.91	25
K01188	Beta-glucosidase	beta-glucosidase	[2]Metabolic pathways; biosynthesis of secondary metabolites; cyanoamino acid metabolism; starch and sucrose metabolism; phenylpropanoid biosynthesis	0.0398	–0.0210	0.90	[1]
K07214	Enterochelin esterase and related enzymes	*fes*	[3]Inorganic ion transport and metabolism	0.0475	–0.1511	0.90	[1]
K03303	Lactate permease	*lctP*	[5]Lactate transmembrane transporter activity	0.0195	–0.0271	0.88	28
K00634	Phosphate butyryltransferase	*ptb*	[2]Metabolic pathways; butanoate metabolism	0.0075	–0.0079	0.85	[1]
K01235	Alpha-glucuronidase	*aguA*	[3]Carbohydrate transport and metabolism	0.0104	–0.0626	0.80	NC
K07561	diphthamide synthase subunit DPH2	*dph2*	[3]Translation, ribosomal structure and biogenesis	0.0030	–0.3881	1.86	01
K01925	UDP-N-acetylmuramoylalanine–D-glutamate ligase	*murD*	[2]Metabolic pathways; D-Glutamine and D-glutamate metabolism; peptidoglycan biosynthesis	0.0620	–0.0857	0.99	[1]
K02437	Glycine cleavage system H protein	*gcvH*	[2]Metabolic pathways; biosynthesis of secondary metabolites; biosynthesis of antibiotics; glycine, serine and threonine metabolism; glyoxylate and dicarboxylate metabolism; carbon metabolism	0.0069	–0.0167	0.93	[1]
K03530	DNA-binding protein HU-beta	*hupB*	[3]DNA binding protein: replication, recombination, and repair	0.0331	–0.0892	0.89	19
K02600	N utilization substance protein A	*nusA*	[3]Transcription	0.1126	–0.0655	0.89	[1]
K02518	Translation initiation factor IF-1	*infA*	[3]Translation, ribosomal structure and biogenesis	0.0346	0.0170	0.88	21

Each column respectively presents information about: 1) KEGG identifier, 2) description of the gene (from KEGG), 3) gene name abbreviation, 4) metabolic pathways in which this gene participates, 5) mean relative abundance of the microbial gene in 42 animals, 6) the partial least squares (PLS) estimate of the regression coefficient using three latent variables, 7) the variable importance in projection (VIP) calculated during the PLS analysis using three latent variables, and 8) the cluster in which the microbial gene was allocated in the final network. 1Microbial genes excluded from the final network due to the 0.80 minimum correlation threshold. NC, Microbial genes not clustered in the final network. Information retrieved from: 2KEGG database, 3NCBI database, 4BioCyc database, and 5UniProt database. The genes in this table explained 63.4% of the variation in FCR (feed conversion ratio). Rows colored in gray correspond to genes simultaneously identified for both FCR and ADG (average daily gain) prediction.

emissions and FCR by Roehe et al. (2016). The trait complexes associated with feed conversion efficiency were further evidenced when analyzing the overlapping genes identified for the prediction of each trait (**Figure 4** and shaded rows in **Tables 2**–**5**), i.e., six microbial genes were identified for the prediction of both FCR and ADG and three genes for the prediction of both RFI and DFI. In agreement, strong correlations were observed for each pair of traits, as shown previously in the literature with the literature (Arthur and Herd, 2008; Herd et al., 2014). These results suggest that different microbial genes can be used to predict each trait. Furthermore, microbial genes overlapping for the prediction of more than one trait might be useful for the interpretation of biological processes explaining the correlation between phenotypes.

TABLE 3 | Summary of microbial genes identified for the prediction of ADG.

KEGG id	Description	Gene name abbreviation	Pathways	Mean abundance	PLS estimate	VIP	Cluster
K01448	N-acetylmuramoyl-L-alanine amidase	*amiABC*	[2]Cationic antimicrobial peptide (CAMP) resistance	0.0236	–0.1937	1.22	06
K00133	Aspartate-semialdehyde dehydrogenase	*asd*	[2]Metabolic pathways; microbial metabolism in diverse environments; biosynthesis of secondary metabolites; biosynthesis of antibiotics; glycine, serine and threonine metabolism; monobactam biosynthesis; cysteine and methionine metabolism; lysine biosynthesis; 2-oxocarboxylic acid metabolism; biosynthesis of amino acids	0.1197	–0.0684	1.20	NC
K01912	Phenylacetate-CoA ligase	*paaK*	[2]Microbial metabolism in diverse environments; phenylalanine metabolism; biofilm formation—*Vibrio cholerae*	0.1543	–0.0980	1.16	16
K02919	Large subunit ribosomal protein L36	*rpmJ*	[2]Ribosome	0.0261	–0.1884	1.04	NC
K02879	Large subunit ribosomal protein L17	*rplQ*	[2]Ribosome	0.0773	0.0746	1.00	21
K02113	F-type H+-transporting ATPase subunit delta	*atpH*	[2]Metabolic pathways; oxidative phosphorylation; photosynthesis	0.0292	–0.0486	1.00	21
K00283	Glycine dehydrogenase subunit 2	*gcvPB*	[2]Metabolic pathways; biosynthesis of secondary metabolites; biosynthesis of antibiotics; glycine, serine and threonine metabolism; glyoxylate and dicarboxylate metabolism; carbon metabolism	0.0284	0.0502	0.99	25
K03775	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD	*slyD*	[5]Posttranslational modification, protein turnover, chaperones	0.0139	0.0672	0.93	22
K07561	Diphthamide synthase subunit DPH2	*dph2*	[5]Translation, ribosomal structure, and biogenesis	0.0030	0.2310	1.20	01
K01925	UDP-N-acetylmuramoylalanine–D-glutamate ligase	*murD*	[2]Metabolic pathways; D-glutamine and D-glutamate metabolism; peptidoglycan biosynthesis	0.0620	0.1155	1.15	[1]
K02437	Glycine cleavage system H protein	*gcvH*	[2]Metabolic pathways; biosynthesis of secondary metabolites; biosynthesis of antibiotics; glycine, serine and threonine metabolism; glyoxylate and dicarboxylate metabolism; carbon metabolism	0.0069	0.1209	1.08	[1]
K03530	DNA-binding protein HU-beta	*hupB*	[5]DNA binding protein: replication, recombination, and repair	0.0331	0.1062	1.07	19
K02600	N utilization substance protein A	*nusA*	[5]Transcription	0.1126	0.0726	1.02	[1]
K02518	Translation initiation factor IF-1	*infA*	[5]Translation, ribosomal structure, and biogenesis	0.0346	0.0646	0.98	21

Each column respectively presents information about: 1) KEGG identifier, 2) description of the gene (from KEGG), 3) gene name abbreviation, 4) metabolic pathways in which this gene participates, 5) mean relative abundance of the microbial gene in 42 animals, 6) the partial least squares (PLS) estimate of the regression coefficient using three latent variables, 7) the variable importance in projection (VIP) calculated during the PLS analysis using three latent variables, and 8) the cluster in which the microbial gene was allocated in the final network. 1Microbial genes excluded from the final network due to the 0.80 minimum correlation threshold. NC, Microbial genes not clustered in the final network. Information retrieved from: 2KEGG database, 3NCBI database, 4BioCyc database, and 5UniProt database. The genes in this table explained 65.4% of the variation in ADG (average daily gain). Rows colored in gray correspond to genes simultaneously identified for both FCR (feed conversion ratio) and ADG prediction.

Metabolic Pathways of Microbial Genes Associated With Efficiency Traits

Our results indicate that most proteins encoded by microbial genes identified for the prediction of FCR were generally involved in carbohydrates metabolism and transport. For example, *aguA* and K01188 are involved in biomass conversion, through the degradation of hemicelluloses and lignocelluloses and lactate biosynthesis (Cairns and Esen, 2010; Lee et al., 2012; Michlmayr and Kneifel, 2014; Li, 2015). Microbial genes *xylE*, *aguA*, and *uidA* are involved in xylan degradation, the main component of hemicellulose (Lee et al., 2012; Fliegerova et al., 2015). Xylose needs to be taken up by a transporter (putatively associated with *xylE*) before it is metabolized, and it has been recognized as a rate-controlling step in bacterial metabolism (Chaillou and Pouwels, 1999). Furthermore, microbial genes such as *uidA* [previously identified by Roehe et al. (2016)], directly involved in carbohydrate metabolism pathways like pentose and glucuronate interconversions and galactose metabolism, are coupled with NAD or NADP oxidoreduction, important for regulating the flux of carbon and energy sources in microorganisms (Spaans et al., 2015). In addition, *punA* (i.e., purine-nucleoside phosphorylase) is involved in the metabolism of nucleotides, nicotinate and nicotinamide (vitamin B3), which also contain NAD and NADP, and is therefore important in carbohydrate, protein, and lipid

TABLE 4 | Summary of microbial genes identified for the prediction of RFI.

KEGG id	Description	Gene name abbreviation	Pathways	Mean abundance	PLS estimate	VIP	Cluster
K03406	Methyl-accepting chemotaxis protein	*mcp*	[2]Two-component system; bacterial chemotaxis	0.0225	–0.0510	1.26	[1]
K03413	Two-component system, chemotaxis family, response regulator CheY	*cheY*	[2]Two-component system; bacterial chemotaxis	0.0018	0.0478	1.16	[1]
K01534	Cd2+/Zn2+-exporting ATPase	*zntA*	[5]Cation-transporting ATPase activity; metal ion binding; nucleotide binding	0.0211	–0.0653	1.16	04
K07258	serine-type D-Ala-D-Ala carboxypeptidase (penicillin-binding protein 5/6)	*dacC*	[2]Metabolic pathways; Peptidoglycan biosynthesis	0.0049	–0.0375	1.14	[1]
K07301	Cation:H+ antiporter	*yrbG*	[3]Inorganic ion transport and metabolism	0.0096	–0.0145	1.09	04
K04720	Threonine-phosphate decarboxylase	*cobD*	[2]Porphyrin and chlorophyll metabolism	0.0034	–0.0501	1.06	04
K03407	Two-component system, chemotaxis family, sensor kinase CheA	*cheA*	[2]Two-component system; bacterial chemotaxis	0.0048	–0.0236	1.04	[1]
K00595	Precorrin-6Y C5,15-methyltransferase (decarboxylating)	*cobL*	[2]Metabolic pathways; porphyrin and chlorophyll metabolism	0.0078	0.0223	1.02	04
K01571	Oxaloacetate decarboxylase, alpha subunit	*oadA*	[2]Metabolic pathways; pyruvate metabolism	0.0165	–0.0501	0.96	04
K02057	Simple sugar transport system permease protein	*ABC.SS.P*	[3]Carbohydrate transport and metabolism	0.0023	–0.1375	0.96	20
K02390	Flagellar hook protein FlgE	*flgE*	[2]Flagellar assembly	0.0015	–0.0376	0.87	[1]
K02417	Flagellar motor switch protein FliN/FliY	*fliN*	[2]Bacterial chemotaxis; flagellar assembly	0.0018	–0.1120	0.77	[1]
K03738	Aldehyde:ferredoxin oxidoreductase	*aor*	[2]Metabolic pathways; Microbial metabolism in diverse environments; Pentose phosphate pathway; Carbon metabolism	0.0144	–0.0657	0.68	NC
K02009	Cobalt transport protein	*cbiN*	[2]ABC transporters	0.0074	–0.1126	0.67	01
K01709	CDP-glucose 4,6-dehydratase	*rfbG*	[2]Metabolic pathways; amino sugar and nucleotide sugar metabolism	0.0041	0.2549	1.46	[1]
K00978	Glucose-1-phosphate cytidylyltransferase	*rfbF*	[2]Metabolic pathways; amino sugar and nucleotide sugar metabolism; starch and sucrose metabolism	0.0042	0.2056	1.23	[1]
K14113	Energy-converting hydrogenase B subunit D	*ehbD*	–	0.0010	0.1703	1.00	NC

Each column respectively presents information about: 1) KEGG identifier, 2) description of the gene (from KEGG), 3) gene name abbreviation, 4) metabolic pathways in which this gene participates, 5) mean relative abundance of the microbial gene in 42 animals, 6) the partial least squares (PLS) estimate of the regression coefficient using three latent variables, 7) the variable importance in projection (VIP) calculated during the PLS analysis using three latent variables, and 8) the cluster in which the microbial gene was allocated in the final network. 1Microbial genes excluded from the final network due to the 0.80 minimum correlation threshold. NC, Microbial genes not clustered in the final network. Information retrieved from: 2KEGG database, 3NCBI database, 4BioCyc database, and 5UniProt database. The genes in this table explained 65.6% of the variation in RFI (residual feed intake). Rows colored in grey correspond to genes simultaneously identified for both RFI and DFI (daily feed intake) prediction.

metabolism reactions. Positive effects of vitamin B3 have been previously observed in healthy rumen microbiomes in beef and dairy cattle (Aschemann et al., 2012; Luo et al., 2017). Microbial genes *uidA* and *punA* were more abundant in efficient animals.

Proteins encoded by *lctP*, K01188, and *ptb*, involved in lactate transport and cellulose and butyrate metabolism, respectively, could be involved in host–microbiome crosstalk mechanisms in cattle due to their participation in metabolic pathways that involve the release of H^+, such as lactate metabolism, potentially reducing microbial fiber-degrading activity and consequently slowing digestion and rumen emptying rate, causing a decrease in appetite (Moran, 2005b). Furthermore, beta-glucosidase is widely present in lactic acid bacteria and is thought to interact with the human host (Michlmayr and Kneifel, 2014). Butyrate has been shown in rats to directly activate the intestinal gluconeogenesis genes in enterocytes *via* an increase in cationic antimicrobial peptides (cAMP, De Vadder et al., 2014). In contrast, *glo1* (more abundant in FCR-H) is involved in methylglyoxal degradation, which is a highly toxic substance that decreases bacterial cell viability, and is produced by bacteria when there is carbohydrate excess and nitrogen limitation (Russell, 1993). Therefore, *glo1* is a strong candidate biomarker of rumen microbiome difference in less efficient animals (i.e., FCR-H).

The microbial gene with highest impact in prediction of ADG was *amiABC*, which is mainly involved in the peptidoglycan turnover through cleavage of glyosidic bonds and release of amino acids and cAMP resistance (Uehara and Park, 2008; Uehara et al., 2010). Some bacteria (mostly pathogenic) have evolved mechanisms of resistance, such as decreased affinity to cAMPs (Anaya-López et al., 2013), and the higher abundance of *amiABC* in animals with lower ADG may be indicative of higher abundance of pathogens, which can cause inflammatory response in the rumen potentially reducing nutrient use and absorption (Reynolds et al., 2017). Brown et al. (2003) demonstrated that acetate and propionate are agonists of the human receptors GPR43 and GPR41, and Hong et al. (2005) proposed that acetate

TABLE 5 | Summary of microbial genes identified for the prediction of DFI.

KEGG id	Description	Gene name abbreviation	Pathways	Mean abundance	PLS estimate	VIP	Cluster
K00370	Nitrate reductase 1, alpha subunit	*narG*	[2]Microbial metabolism in diverse environments; nitrogen metabolism; two-component system	0.0022	–0.2272	1.22	[1]
K01858	Myo-inositol-1-phosphate synthase	*INO1*	[2]Metabolic pathways; biosynthesis of antibiotics; streptomycin biosynthesis; inositol phosphate metabolism	0.0542	–0.0459	1.14	[1]
K03685	Ribonuclease III	*rnc*	[2]Ribosome biogenesis in eukaryotes; proteoglycans in cancer	0.0288	–0.0097	1.13	[1]
K00613	Glycine amidinotransferase	*GATM*	[2]Metabolic pathways; glycine, serine and threonine metabolism; arginine and proline metabolism	0.0019	–0.1417	1.09	[1]
K02428	XTP/dITP diphosphohydrolase	*rdgB*	[2]Metabolic pathways; purine metabolism	0.0147	–0.0216	0.94	02
K03602	Exodeoxyribonuclease VII small subunit	*xseB*	[2]Mismatch repair	0.0035	0.0803	0.94	02
K03210	Preprotein translocase subunit YajC	*yajC*	[2]Bacterial secretion system; quorum sensing; protein export	0.0069	0.1317	0.93	[1]
K12340	Outer membrane channel protein TolC	*tolC*	[2]Beta-lactam resistance; cationic antimicrobial peptide (CAMP) resistance; two-component system; bacterial secretion system; plant–pathogen interaction; pertussis	0.0157	0.0068	0.92	02
K03043	DNA-directed RNA polymerase subunit beta	*rpoB*	[2]Metabolic pathways; purine metabolism; pyrimidine metabolism; RNA polymerase	1.2470	–0.0995	0.91	NC
K04751	Nitrogen regulatory protein P-II 1	*glnB*	[2]Two-component system	0.0151	0.0613	0.91	02
K03625	N utilization substance protein B	*nusB*	[3]Transcription termination	0.0135	0.0766	0.91	02
K06178	Ribosomal large subunit pseudouridine synthase B	*rluB*	[3]Translation, ribosomal structure, and biogenesis	0.0693	–0.0038	0.85	02
K05349	Beta-glucosidase	*bglX*	[2]Metabolic pathways; biosynthesis of secondary metabolites; cyanoamino acid metabolism; starch and sucrose metabolism; phenylpropanoid biosynthesis	0.2272	0.0063	0.84	[1]
K05515	Penicillin-binding protein 2	*mrdA*	[2]Peptidoglycan biosynthesis; beta-lactam resistance	0.0295	0.0214	0.82	02
K04764	Integration host factor subunit alpha	*ihfA*	[3]DNA binding: replication, recombination, and repair	0.0041	0.0306	0.80	02
K01709	CDP-glucose 4,6-dehydratase	*rfbG*	[2]Metabolic pathways; amino sugar and nucleotide sugar metabolism	0.0041	0.2412	1.53	[1]
K00978	Glucose-1-phosphate cytidylyltransferase	*rfbF*	[2]Metabolic pathways; amino sugar and nucleotide sugar metabolism; starch and sucrose metabolism	0.0042	0.2634	1.43	[1]
K14113	Energy-converting hydrogenase B subunit D	*ehbD*	–	0.0010	0.1594	1.16	NC

Each column respectively presents information about: 1) KEGG identifier, 2) description of the gene (from KEGG), 3) gene name abbreviation, 4) metabolic pathways in which this gene participates, 5) mean relative abundance of the microbial gene in 42 animals, 6) the partial least squares (PLS) estimate of the regression coefficient using three latent variables, 7) the variable importance in projection (VIP) calculated during the PLS analysis using three latent variables, and 8) the cluster in which the microbial gene was allocated in the final network. 1Microbial genes excluded from the final network due to the 0.80 minimum correlation threshold. NC, Microbial genes not clustered in the final network. Information retrieved from: 2KEGG database, 3NCBI database, 4BioCyc database, and 5UniProt database. The genes in this table explained 72.9% of the variation in DFI (daily feed intake). Rows colored in gray correspond to genes simultaneously identified for both RFI (residual feed intake) and DFI prediction.

and propionate induce lipid accumulation and inhibition of lipolysis through the GPR43 receptor in mice. These genes are also part of the bovine genome, where they mediate an inhibitory effect of acetate, propionate, and butyrate on cAMP signaling (Wang et al., 2009). This could indicate that, in less efficient animals (lower ADG), the lower amount of acetate, propionate, and butyrate may lead to decreased inhibition of lipolysis by the host, which potentially results in lower ADG. Alternatively, the lower amount of VFAs in these animals may lead to decreased inhibition of cAMP signaling and increased release of cAMPs by the host to the rumen. The cAMPs act primarily on organisms without effective resistance mechanisms, consequently increasing the relative abundance of cAMP-resisting organisms and of the microbial genes encoding for the resistance. Two other microbial genes identified in the present research are part of the cAMP resistance pathway—*lpxA* and *tolC* (associated with FCR and DFI, respectively). Although all three genes (*amiABC*, *lpxA*, and *tolC*) are part of the same pathway, they present opposite tendencies—while *lpxA* and *tolC* are proportionally highly abundant in animals with higher ADG and lower FCR, *amiABC* is relatively highly abundant in animals with lower ADG and higher FCR. The gene *lpxA* is related to lipid A integration in the cell wall, as a preventive measure against the hosts' immune system, and *tolC* is involved in the efflux of antibiotics (Raetz et al., 2007; Zgurskaya et al., 2011). This could be indicative of the different cAMP resistance mechanisms evolved by bacterial

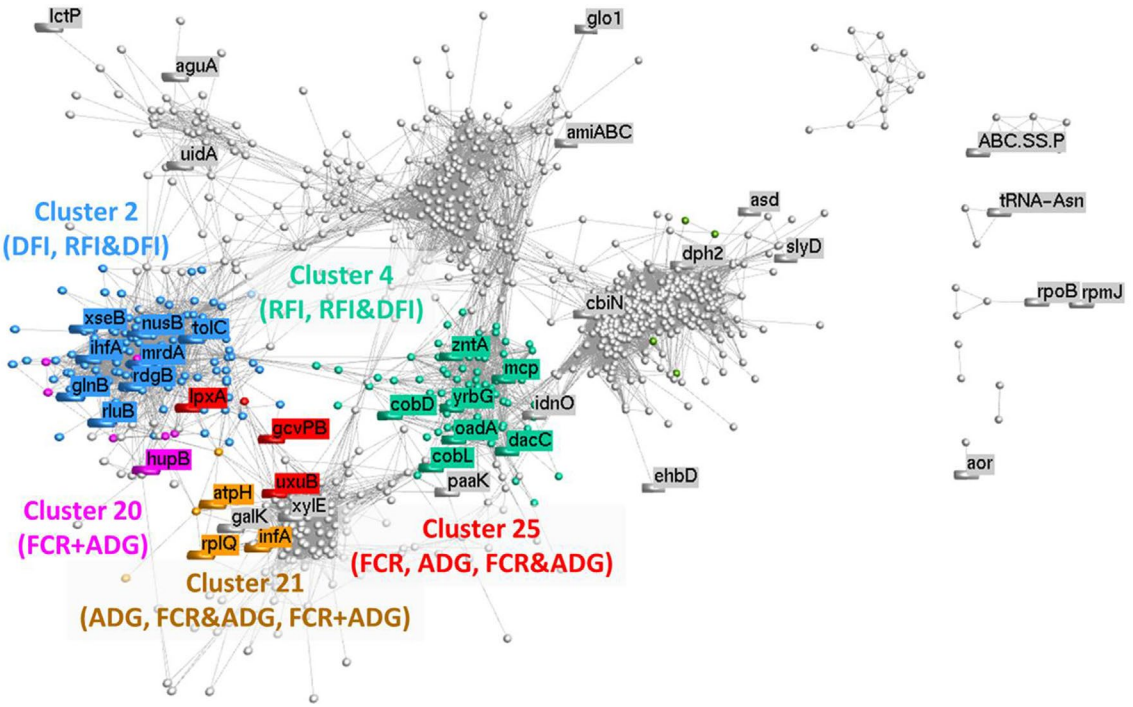


FIGURE 5 | Correlation network analysis of metagenomic data: Each node represents a vector of relative abundances of each microbial gene in all 42 animals, and the edges represent a correlation between the microbial genes. A minimum correlation threshold of 0.80 was applied to the network. Different colors illustrate different clusters, which were calculated using MCL method (inflation: 2; preinflation: 2; scheme: 6). Clusters identified by numbers were found to be significantly ($P < 0.05$) enriched for microbial genes identified for the traits whose abbreviations are between brackets (FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake; FCR&ADG, set including microbial genes identified for prediction of either FCR and/or ADG; RFI&DFI, set including microbial genes identified for prediction of RFI and/or DFI; FCR+ADG, set including microbial genes simultaneously identified for prediction of both traits FCR and ADG).

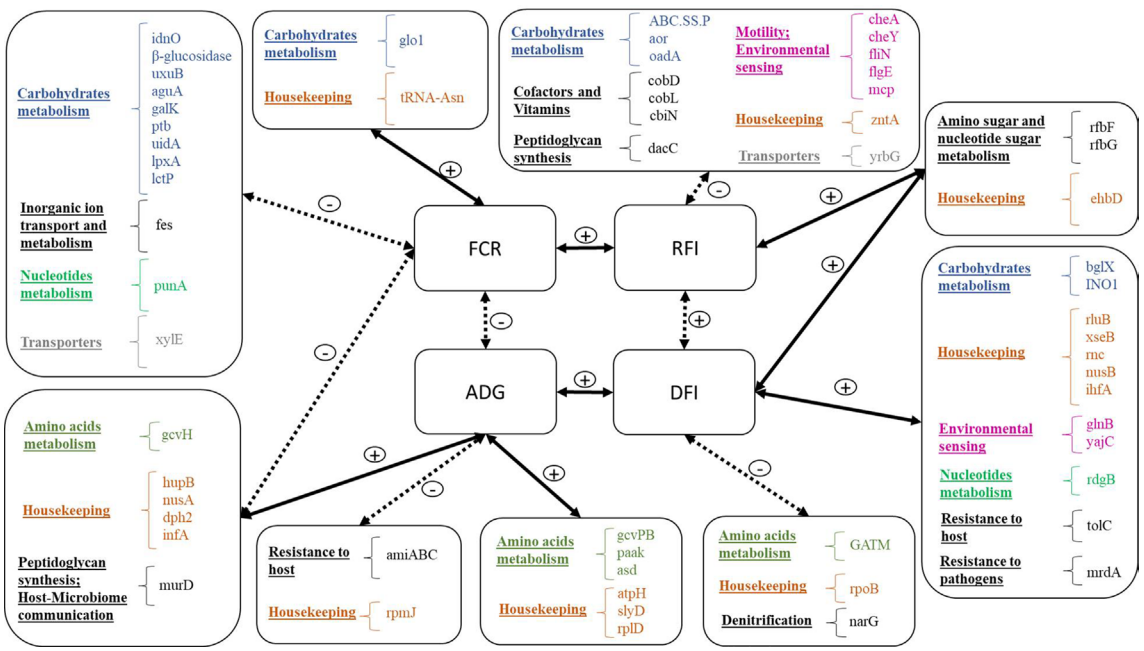


FIGURE 6 | Summary of microbial genes identified for the prediction of each trait: Traits are located in the four central boxes: FCR, feed conversion ratio; ADG, average daily gain; RFI, residual feed intake; DFI, daily feed intake. Solid lines represent positive correlations, and dotted lines represent negative correlations. Microbial genes are listed in the outside boxes, organized by general function, and each general function is represented by a different color.

organisms, which include modification of the cell external surface, efflux pumps, and biosynthesis and crosslinking of cell envelope components (Nizet, 2006).

The set of microbial genes associated with ADG included mostly housekeeping genes and genes related to amino acid metabolism and transport. Artegoitia et al. (2017) found a link between ruminal aromatic amino acids synthesis such as phenylalanine and high ADG in beef steers. For example, *paak* [previously mentioned by Kamke et al. (2016) related to sheep with high production of methane] and *asd* encode proteins that respectively catalyze phenylalanine and phenylacetate (related to aspartate degradation and biosynthesis of amino acids including threonine), with release of H^+. In the current research, both of these genes were positively correlated to ADG, which is supported by the positive correlations between ADG and dry matter intake (DMI), between DMI and methane emissions, and between methane emissions and body weight measurements (weaning weight, yearling weight, and final weight), previously observed in cattle (Koots et al., 1994; Arthur et al., 2001; Herd et al., 2014).

Some housekeeping genes were simultaneously identified for the prediction of FCR and ADG, such as protein translation from diphthamide (*dph2*) or peptidoglycan biosynthesis (*murD*), both more abundant in efficient animals (higher ADG and lower FCR). The importance of diphthamide biosynthesis in archaea is not yet fully known (Narrowe et al., 2018). Microbial gene *murD* is related to the glutamate–glutamine cycle, an important appetite regulator in humans (Delgado, 2013), but in the present research, it was not associated to DFI.

Proteins encoded by microbial genes associated with RFI are mostly related to chemotaxis (*cheA* and *cheY*), detoxification (Cd2+/Zn2+-exporting ATPase, *zntA*), and vitamin B12 production (*cbiN*, *cobD*, and *cobL*). The negative correlation of microbial genes involved in chemotaxis and motility with RFI may suggest an increased microbial metabolism in efficient animals, derived from their ability to sense chemical gradients in their surrounding environment and to react accordingly, i.e., moving closer to nutrients (Rajagopala et al., 2007). Microbial gene *zntA* was also more abundant in efficient animals and plays a role in the homeostasis of transition metals (Cd^{2+}, Zn^{2+}), participating in functional pathways ranging from cellular respiration to gene expression (Fraústro da Silva and Williams, 2001). Finally, higher relative abundance of microbial genes involved in vitamin B12 production (*cbiN*, *cobD*, and *cobL*) was observed in more efficient animals. This essential cofactor needs to be taken up directly from the diet or to be made available for animal absorption by the rumen microbial organisms because it is not produced by eukaryotes (Warren et al., 2002). Furthermore, vitamin B12 has been previously associated with increased cobalt content on high-fiber diets and increased VFA, such as acetate (Beaudet et al., 2017), which may affect the animals' appetite (Frost et al., 2014), in line with our observation of higher relative abundance of these genes in more efficient animals, i.e., animals with lower feed intake than expected.

The four most important microbial genes identified for the prediction of DFI included the three microbial genes also identified for prediction of RFI (*rfbG*, *rfbF*, and *ehbD*) and *narG*. Microbial genes *rfbG* and *rfbF* (VIP > 1.4) are part of the *rfc* region (Morona et al., 1994) and are related to nucleotide sugar metabolism, which is necessary for the production of microbial lipopolysaccharide (LPS). LPS is a major virulence factor of Gram-negative bacteria, particularly due to the O-antigen, paramount for host colonization and niche adaptation by bacterial organisms, due to its part in the protection from host immune response (Reeves, 1995; Samuel and Reeves, 2003; Geue et al., 2017). Both genes *rfbG* and *rfbF* showed a positive correlation to RFI and DFI, supporting our hypothesis that the use of energy to stimulate the innate immune system against pathogens increases DFI and reduces feed conversion efficiency as determined by RFI (Neal et al., 1991; Jing et al., 2014; Vigors et al., 2016). Other microbial genes positively correlated to DFI were found to be involved in resistance mechanisms, such as the penicillin-binding protein 2-encoding gene (*mrdA*), which belongs to the peptidoglycan and beta-lactam resistance metabolic pathways. These proteins are transpeptidases or carbopeptidases involved in peptidoglycan metabolism and have an important role against beta-lactam resistance (Zapun et al., 2008). The microbial gene myo-inositol-1-phosphate synthase (*INO1*) is related to antibiotic biosynthesis, including streptomycin. Microbial gene *ehbD* is a subunit of the energy-converting hydrogenase B, found in methanogens such as *Methanococcus maripaludis*. This microbial gene is important due to its role in autotrophic CO_2 assimilation (Porat et al., 2006), having implications for microbial growth. Furthermore, *narG*, part of the *narGHIJ* operon, essential for some microorganisms to gather energy under anaerobic conditions by the reduction in nitrate to nitrite in a denitrification process (Blasco et al., 1990; Latham et al., 2016), was proportionally more abundant in animals with low DFI.

The microbial gene *nusB* (associated with DFI) is part of a set of *nus* genes, which also includes *nusA* (identified for prediction of FCR and ADG). Genes in the *nus* complex are involved in transcription termination and antitermination processes, such as Rho-dependent transcriptional termination (Torres et al., 2004), which is the regulatory mechanism involved in the efficient transcription of the tryptophan operon (Farnham et al., 1982; Kuroki et al., 1982; Prasch et al., 2009). The *nus*-complex microbial genes were found to be relatively more abundant in efficient animals. This association may be due to the influence of the *nus* genes, which extends from the ribosomal operons to the tryptophan operon and constitutes a good example of how termination and antitermination processes can control gene expression, occurring during RNA transcription, and potentially positively impacting bacterial growth and rumen fermentation processes.

Although microbial genes *amiABC*, *tolC*, *glo1, rfbF*, *rfbG*, *and lpxA* were identified in the present research for the prediction of different traits, all are associated with bacterial defense mechanisms either from other bacteria or from the host. The majority of these genes had higher abundance in less efficient animals. This suggests that the presence of either bacterial pathogens in the rumen or antibiotics produced as host immune responses might represent a significant energy sink, impairing feed conversion efficiency.

Further improvement of prediction of feed conversion traits using metagenomic information may be achieved through

the integration of protein, enzyme, and pathway data from the Hungate collection (Seshadri et al., 2018) and the large rumen metagenomic reference dataset (Stewart et al., 2018).

CONCLUSIONS

The results presented here suggest that relative abundances of rumen microbial genes may be highly informative predictors of feed conversion efficiency, growth rate, and feed intake, which are labor intensive, time consuming, and expensive traits to record. Most microbial genes identified for the prediction of traits in this research were trait specific. Microbial genes related to cellulose and hemicellulose degradation, vitamin B12 synthesis, and amino acids metabolism were associated to enhanced feed conversion efficiency (FCR or RFI), while those involved in nucleotide sugars metabolism, pathogen LPS synthesis, cAMP resistance, and degradation of toxic compounds were associated with inefficient feed conversion. Furthermore, we identified specific microbial genes encoding proteins related to the crosstalk between the microbiome and the host cells, such as *murD* and *amiABC*, and associated to gene expression regulatory mechanisms, such as *nusA* and *nusB*. Thus, our results provide a deeper understanding of the potential influence of the rumen microbiome on the feed conversion efficiency of its host, highlighting specific enzymes involved in metabolic pathways that reflect the complex functional networks impacting the conversion of feed into animal products such as meat.

AUTHOR CONTRIBUTIONS

RR and MW conceived and designed the overall study, and JL, MA, TF, and RR conceived, designed, and executed the bioinformatics analysis. RS and MW carried out the bioinformatics to obtain the rumen microbial gene abundances. C-AD, TS, RD, and AW provided essential insight into feed conversion efficiency, rumen metabolism, nutrition, and microbiology. JL, MA, and RR wrote the initial draft, and subsequently, all authors contributed intellectually to the interpretation and presentation of the results.

FUNDING

The project was supported by grants from the Biotechnology and Biological Sciences Research Council (BBSRC BB/N01720X/1 and BB/N016742/1) and by the Scottish Government (RESAS Division) as part of the 2016–2021 commission. The research is based on data from experiments funded by the Scottish Government as part of the 2011–2016 commission, Agriculture and Horticulture Development Board (AHDB) Beef & Lamb, Quality Meat Scotland (QMS), and Department for Environment Food & Rural Affairs (Defra).

ACKNOWLEDGMENTS

We thank Dr Irene Cabeza Luna, Laura Nicoll, Lesley Deans, and Claire Broadbent for the excellent technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00701/full#supplementary-material

REFERENCES

Anaya-López, J. L., López-Meza, J. E., and Ochoa-Zarzosa, A. (2013). Bacterial resistance to cationic antimicrobial peptides. *Crit. Rev. Microbiol.* 39, 180–195. doi: 10.3109/1040841X.2012.699025

Arora, T., Sharma, R., and Frost, G. (2011). Propionate. Anti-obesity and satiety enhancing factor? *Appetite* 56, 511–515. doi: 10.1016/j.appet.2011.01.016

Artegoitia, V. M., Foote, A. P., Lewis, R. M., and Freetly, H. C. (2017). Rumen fluid metabolomics analysis associated with feed efficiency on crossbred steers. *Sci. Rep.* 7, 1–14. doi: 10.1038/s41598-017-02856-0

Arthur, J. P. F., and Herd, R. M. (2008). Residual feed intake in beef cattle. *Rev. Bras. Zootec.* 37, 269–279. doi: 10.1590/S1516-35982008001300031

Arthur, P. F., Renand, G., and Krauss, D. (2001). Genetic and phenotypic relationships among different measures of growth and feed efficiency in young Charolais bulls. *Livest. Prod. Sci.* 68, 131–139. doi: 10.1016/S0301-6226(00)00243-8

Aschemann, M., Lebzien, P., Hüther, L., Südekum, K.-H., and Dänicke, S. (2012). Effect of niacin supplementation on rumen fermentation characteristics and nutrient flow at the duodenum in lactating dairy cows fed a diet with a negative rumen nitrogen balance. *Arch. xAnim. Nutr.* 66, 303–318. doi: 10.1080/1745039x.2012.697353

Auffret, M. D., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., John Wallace, R., Freeman, T. C., et al. (2017). The rumen microbiome as a reservoir of antimicrobial resistance and pathogenicity genes is directly affected by diet in beef cattle. *Microbiome* 5, 159–169. doi: 10.1186/s40168-017-0378-z

Bateman, A. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049

Beaudet, V., Gervais, R., Graulet, B., Nozière, P., Doreau, M., Fanchone, A., et al. (2017). Effects of dietary nitrogen levels and carbohydrate sources on apparent ruminal synthesis of some B vitamins in dairy cows. *J. Dairy Sci.* 99, 2730–2739. doi: 10.3168/jds.2017-100-1-0849

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., 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.* 107, 18933–18938. doi: 10.1073/pnas.1007028107

Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* 70, 567–590. doi: 10.1152/physrev.1990.70.2.567

Berry, D. P., and Crowley, J. J. (2013). Cell biology symposium: genetics of feed efficiency in dairy and beef cattle. *J. Anim. Sci.* 91, 1594–1613. doi: 10.2527/jas.2012-5862

Blasco, F., Iobbi, C., Ratouchniak, J., Bonnefoy, V., and Chippaux, M. (1990). Nitrate reductases of *Escherichia coli*: sequence of the second nitrate reductase and comparison with that encoded by the narGHJI operon. *MGG Mol. Gen. Genet.* 222, 104–111.

Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L., Daniels, D., et al. (2003). The orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J. Biol. Chem.* 278, 11312–11319. doi: 10.1074/jbc.m211609200

Brulc, J. M., Antonopoulos, D. A., Miller, M. E. B., Wilson, M. K., Yannarell, A. C., Dinsdale, E. A., et al. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci.* 106, 1948–1953. doi: 10.1073/pnas.0806191105

Cairns, J. R. K., and Esen, A. (2010). β-Glucosidases. *Cell. Mol. Life Sci.* 67, 3389–3405. doi: 10.1007/s00018-010-0399-2

Chaillou, P., and Pouwels, P. H. (1999). Transport of D-xylose in *Lactobacillus pentosus*, *Lactobacillus casei*, and *Lactobacillus plantarum*: evidence for a mechanism of facilitated diffusion *via* the phosphoenolpyruvate: mannose phosphotransferase system. *Society* 181, 4768–4773.

Cherbut, C. (2003). Motor effects of short-chain fatty acids and lactate in the gastrointestinal tract. *Proc. Nutr. Soc.* 62, 95–99. doi: 10.1079/PNS2002213

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

De Vadder, F., Kovatcheva-Datchary, P., Goncalves, D., Vinera, J., Zitoun, C., Duchampt, A., et al. (2014). Microbiota-generated metabolites promote metabolic benefits *via* gut–brain neural circuits. *Cell* 156, 84–96. doi: 10.1016/j.cell.2013.12.016

Delgado, T. C. (2013). Glutamate and GABA in appetite regulation. *Front. Endocrinol.* 4, 103. doi: 10.3389/fendo.2013.00103

Duthie, C.-A., Rooke, J. A., Troy, S., Hyslop, J. J., Ross, D. W., Waterhouse, A., et al. (2015). Impact of adding nitrate or increasing the lipid content of two contrasting diets on blood methaemoglobin and performance of two breeds of finishing beef steers. *Animal* 10, 786–795. doi: 10.1017/S1751731115002657

Duthie, C. A., Troy, S. M., Hyslop, J. J., Ross, D. W., Roehe, R., and Rooke, J. A. (2017). The effect of dietary addition of nitrate or increase in lipid concentrations, alone or in combination, on performance and methane emissions of beef cattle. *Animal* 12, 280–287. doi: 10.1017/S175173111700146X

Eurostat. (2018). Eurostat online table Bovine, alive, EU-28. Luxembourg: Eurostat. Available at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=apro_mt_lscatl&lang=en (Accessed October 16, 2018).

Farnham, P. J., Greenblatt, J., and Platt, T. (1982). Effects of NusA protein on transcription termination in the tryptophan operon of *Escherichia coli. Cell* 29, 945–951. doi: 10.1016/0092-8674(82)90457-3

Fliegerova, K., Kaerger, K., Kirk, P., and Voigt, K. (2015). "7. Rumen fungi," in *Rumen Microbiology: From Evolution to Revolution* (India: Springer India), 97–112. doi: 10.1007/978-81-322-2401-3

Fraústro da Silva, J., and Williams, R. (2001). *The biological chemistry of the elements: the inorganic chemistry of life*. New York: Oxford University Press.

Freeman, T. C., Goldovsky, L., Brosch, M., van Dongen, S., Maziere, P., Grocock, R. J., et al. (2007). Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Comput. Biol.* 3, e206 (11 pages). doi: 10.1371/journal.pcbi.0030206

Frost, G., Sleeth, M. L., Sahuri-Arisoylu, M., Lizarbe, B., Cerdan, S., Brody, L., et al. (2014). The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nat. Commun.* 5, 3611. doi: 10.1038/ncomms4611

Gerber, P. J., Steinfeld, H., Henderson, B., Mottet, A., Opio, C., Dijkman, J., et al. (2013). Tackling climate change through livestock - A global assessment of emissions and mitigation opportunities. Rome: Food and Agriculture Organization of the United Nations (FAO).

Geue, L., Menge, C., Eichhorn, I., Semmler, T., Wieler, L. H., Pickard, D., et al. (2017). Evidence for contemporary switching of the O-antigen gene cluster between shiga toxin-producing escherichia coli strains colonizing cattle. *Front. Microbiol.* 8, 424. doi: 10.3389/fmicb.2017.00424

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F. et al. (2010). Food security: the challenge of feeding 9 billion people. *Science* 327, 812–818. doi: 10.1126/science.1185383

Guan, L. L., Nkrumah, J. D., Basarab, J. A., and Moore, S. S. (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol. Lett.* 288, 85–91. doi: 10.1111/j.1574-6968.2008.01343.x

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Global Rumen Census Collaborators, G., et al. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* 5, 14567. doi: 10.1038/srep14567

Herd, R. M., Arthur, P. F., Donoghue, K. A., Bird, S. H., Bird-Gardiner, T., and Hegarty, R. S. (2014). Measures of methane production and their phenotypic relationships with dry matter intake, growth, and body composition traits in beef cattle. *J. Anim. Sci.* 92, 5267–5274. doi: 10.2527/jas.2014-8273

Hernandez-Sanabria, E., Goonewardene, L. A., Wang, Z., and Durunna, O. N. (2011). Impact of feed efficiency and diet on the adaptive variations in the bacterial community in the rumen fluid of cattle. *Appl. Environ. Microbiol.* 1, 1203–1214. doi: 10.1128/AEM.05114-11

Hong, Y.-H., Nishimura, Y., Hishikawa, D., Tsuzuki, H., Miyahara, H., Gotoh, C., et al. (2005). Acetate and propionate short chain fatty acids stimulate adipogenesis via GPCR43. *Endocrinology* 146, 5092–5099. doi: 10.1210/en.2005-0545

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

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

Jiang, X., Neal, B., Santiago, F., Lee, S. J., and Reeves, P. R. (2006). Structure and sequence of the rfb (O antigen) gene cluster of *Salmonella serovar typhimuriunfi* (strain LT2). *Mol. Microbiol.* 5, 695–713. doi: 10.1111/j.1365-2958.1991.tb00741.x

Jing, L., Zhang, R., Liu, Y., Zhu, W., and Mao, S. (2014). Intravenous lipopolysaccharide challenge alters ruminal bacterial microbiota and disrupts ruminal metabolism in dairy cattle. *Br. J. Nutr.* 112, 170–182. doi: 10.1017/S000711451400066X

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. doi: 10.1186/s40168-016-0201-2

Kanehisa Laboratories (2018). KEGG MAPPER. Available at: https://www.genome.jp/kegg/tool/map_pathway1.html [Accessed October 1, 2018]

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1016/j.meegid.2016.07.022

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., et al. (2017). The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* 201, 1–9. doi: 10.1093/bib/bbx085

Koots, K. R., Gibson, J. P., and Wilton, J. W. (1994). Analyses of published genetic parameter estimates for beef production traits. 2. Phenotypic and genetic correlations. *Anim. Breed. Abstr.* 62, 825–853. Available at: http://agris.fao.org/agris-search/search.do?recordID=GB9508564

Kuroki, K., Ishii, S., Kano, Y., Miyashita, T., Nishi, K., and Imamoto, F. (1982). Involvement of the nusA and nusB gene products in transcription of *Escherichia coli* tryptophan operon *in vitro. MGG Mol. Gen. Genet.* 185, 369–371. doi: 10.1007/BF00330816

Latham, E. A., Anderson, R. C., Pinchak, W. E., and Nisbet, D. J. (2016). Insights on alterations to the rumen ecosystem by nitrate and nitrocompounds. *Front. Microbiol.* 7, 228. doi: 10.3389/fmicb.2016.00228

Lee, C. C., Kibblewhite, R. E., Wagschal, K., Li, R., and Orts, W. J. (2012). Isolation of α-glucuronidase enzyme from a rumen metagenomic library. *Protein J.* 31, 206–211. doi: 10.1007/s10930-012-9391-z

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

Li, R. W. (2015). "16. Rumen metagenomics," in *Rumen Microbiology: From Evolution to Revolution* (India: Springer India), 223–245. doi: 10.1007/978-81-322-2401-3

Luo, D., Gao, Y., Lu, Y., Qu, M., Xiong, X., Xu, L., et al. (2017). Niacin alters the ruminal microbial composition of cattle under high-concentrate condition. *Anim. Nutr.* 3, 180–185. doi: 10.1016/j.aninu.2017.04.005

Lyons, T., Bielak, A., Doyle, E., and Kuhla, B. (2018). Variations in methane yield and microbial community profiles in the rumen of dairy cows as they pass through stages of first lactation. *J. Dairy Sci.* 101, 5102–5114. doi: 10.3168/jds.2017-14200

Maldini, G., and Allen, M. S. (2018). Temporal effects of ruminal propionic acid infusion on feeding behavior of Holstein cows in the postpartum period. *J. Dairy Sci.* 101, 3077–3084 doi: 10.3168/jds.2017-13857

Meale, S. J., Auffret, M. D., Watson, M., Morgavi, D. P., Cantalapiedra-Hijar, G., Duthie, C. A., et al. (2018). Fat accretion measurements strengthen the relationship between feed conversion efficiency and nitrogen isotopic discrimination while rumen microbial genes contribute little. *Sci. Rep.* 8, 3854. doi: 10.1038/s41598-018-22103-4

Michlmayr, H., and Kneifel, W. (2014). β-Glucosidase activities of lactic acid bacteria: mechanisms, impact on fermented food and human health. *FEMS Microbiol. Lett.* 352, 1–10. doi: 10.1111/1574-6968.12348

Moran, J. (2005a). "Chapter 5: How the rumen works," in *Tropical dairy farming: feeding management for small holder dairy farmers in the humid tropics*. (Collingwood: Csiro Publishing), 41–49. doi: 10.1016/0022-1694(84)90033-7

Moran, J. (2005b). "Chapter 13. Problems with unbalanced diets," in *Tropical Dairy Farming: Feeding Management for Small Holder Dairy Farmers in the Humid Tropics.* doi: 10.1071/9780643093133

Morona, R., Mavris, M., Fallarino, A., and Manning, P. A. (1994). Characterization of the rfc region of *Shigella flexneri*. *J. Bacteriol.* 176, 733–747. doi: 10.1128/jb.176.3.733-747.1994

Myer, P. R., Smith, T. P. L., Wells, J. E., Kuehn, L. A., and Freetly, H. C. (2015). Rumen microbiome from steers differing in feed efficiency. *PLoS One* 10, e0129174. doi: 10.1371/journal.pone.0129174

Narrowe, A. B., Spang, A., Stairs, C. W., Caceres, E. F., Baker, B. J., Miller, C. S., et al. (2018). Complex evolutionary history of translation elongation factor 2 and diphthamide biosynthesis in archaea and parabasalids. *Genome Biol. Evol.* 10, 2380–2393. doi: 10.1093/gbe/evy154

Neal, B., Jiang, X.-M., Lee, S. J., Santiago, F., Reeves, P. R., and Romana, L. K. (1991). Structure and sequence of the rfb (O antigen) gene cluster of Salmonella serovar typhimurium (strain LT2). *Mol. Microbiol.* 5, 695–713. doi: 10.1111/j.1365-2958.1991.tb00741.x

Nielsen, M. K., MacNeil, M. D., Dekkers, J. C. M., Crews, D. H., Jr., Rathje, T. A., Enns, R. M., et al. (2013). Review: Life-cycle, total-industry genetic improvement of feed efficiency in beef cattle: Blueprint for the Beef Improvement Federation. *Prof. Anim.* Sci. 29, 559–565. doi: 10.15232/s1080-7446(15)30285-0

Nizet, V. (2006). Antimicrobial peptide resistance mechanisms of human bacterial pathogens. *Curr Issues Mol Biol* 8, 11–26 Available at: https://www.ncbi.nlm.nih.gov/pubmed/16450883.

Oba, M., and Allen, M. S. (2003). Intraruminal infusion of propionate alters feeding behavior and decreases energy intake of lactating dairy cows. *J. Nutr* 133, 1094–1099. doi: 10.3168/jds.S0022-0302(03)73889-2

Porat, I., Kim, W., Hendrickson, E. L., Xia, Q., Zhang, Y., Wang, T., et al. (2006). Disruption of the operon encoding Ehb hydrogenase limits anabolic CO 2 assimilation in the archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 188, 1373–1380. doi: 10.1128/JB.188.4.1373-1380.2006

Prasch, S., Jurk, M., Washburn, R. S., Gottesman, M. E., Wöhrl, B. M., and Rösch, P. (2009). RNA-binding specificity of *E. coli* NusA. *Nucleic Acids Res.* 37, 4736–4742. doi: 10.1093/nar/gkp452

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, S., Manichanh, C., et al., (2010). A human gut microbial gene catalog established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821

Raetz, C. R. H., Reynolds, C. M., Trent, M. S., and Bishop, R. E. (2007). Lipid A modification systems in Gram-negative bacteria. *NIH Public Access* 76, 295–329. doi: 10.1146/annurev.biochem.76.010307.145803

Rajagopala, S. V., Titz, B., Goll, J., Parrish, J. R., Wohlbold, K., McKevitt, M. T., et al. (2007). The protein network of bacterial motility. *Mol. Syst. Biol.* 3, 128. doi: 10.1038/msb4100166

Reeves, P. (1995). Role of O-antigen variation in the immune response. *Trends Microbiol.* 3, 381–386. doi: 10.1016/S0966-842X(00)88983-0

Reynolds, J. G., Foote, A. P., Freetly, H. C., Oliver, W. T., and Lindholm-Perry, A. K. (2017). Relationships between inflammation- and immunity-related transcript abundance in the rumen and jejunum of beef steers with divergent average daily gain. *Anim. Genet.* 48, 447–449. doi: 10.1111/age.12546

Roehe, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., McKain, N., Ross, D. W., et al. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* 12, e1005846. doi: 10.1371/journal.pgen.1005846

Rooke, J. A., Wallace, R. J., Duthie, C.-A., McKain, N., Souza, S. M. de, Hyslop, J. J., et al. (2014). Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype. *Br. J. Nutr.* 112, 398–407. doi: 10.1017/s0007114514000932

Ross, E. M., Moate, P. J., Marett, L. C., Cocks, B. G., and Hayes, B. J. (2013). Metagenomic predictions: from microbiome to complex health and environmental phenotypes in humans and cattle. *PLoS One* 8, e73056. doi: 10.1371/journal.pone.0073056

Russell, J. B. (1993). Glucose toxicity in *Prevotella ruminicola*: methylglyoxal accumulation and its effect on membrane physiology. *Appl. Environ. Microbiol.* 59, 2844–2850. Available at: https://www.ncbi.nlm.nih.gov/pubmed/8215358

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

Russell, J. B., and Hespell, R. B. (1981). Microbial rumen fermentation. *J. Dairy Sci.* 64, 1153–1169. doi: 10.3168/jds.S0022-0302(81)82694-X

Sakata, T., and Tamate, H. (1979). Rumen epithelial cell proliferation accelerated by propionate and acetate. *J. Dairy Sci.* 61, 1109–1113. doi: 10.3168/jds.S0022-0302(78)83694-7

Samuel, G., and Reeves, P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr. Res.* 338, 2503–2519. doi: 10.1016/j.carres.2003.07.009

Sasson, G., Ben-Shabat, S., Seroussi, E., Doron-Faigenboim, A., Shterzer, N., Yaacoby, S., et al. (2017). Heritable bovine rumen bacteria are phylogenetically related and correlated with the cow's capacity to harvest. *Am. Soc. Microbiol.* 8, e00703–17. doi: 10.1128/mBio.00703-17

Scollan, N. D., Morrisson, S., Popova, M., Attwood, G. T., Creevey, C. J., Tapio, I., et al. (2018). Addressing global ruminant agricultural challenges through understanding the rumen microbiome: past, present, and future. *Front. Microbiol.* 9, 2161. doi: 10.3389/fmicb.2018.02161

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

Shabat, S., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., Berg Miller, M. E., et al. (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J.* 10, 2958–2972. doi: 10.1038/ismej.2016.62

Spaans, S. K., Weusthuis, R. A., van der Oost, J., and Kengen, S. W. M. (2015). NADPH-generating systems in bacteria and archaea. *Front. Microbiol.* 6, 742. doi: 10.3389/fmicb.2015.00742

Stewart, R. D., Auffret, M. D., Warr, A., Wiser, A. H., Press, M. O., Langford, K. W., et al. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* 9, 870. doi:10.1038/s41467-018-03317-6

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J. P., et al. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* 11, 2574–2584. doi: 10.1111/j.1462-2920.2009.01982.x

Torres, M., Balada, J. M., Zellars, M., Squires, C., and Squires, C. L. (2004). *In vivo* effect of NusB and NusG on rRNA transcription antitermination. *J. Bacteriol.* 186, 1304–1310. doi: 10.1128/JB.186.5.1304-1310.2004

Troy, S. M., Duthie, C. A., Hyslop, J. J., Roehe, R., Ross, D. W., Wallace, R. J., et al. (2015). Effectiveness of nitrate addition and increased oil content as methane mitigation strategies for beef cattle fed two contrasting basal diets. *J. Anim. Sci.* 93, 1815–1823. doi: 10.2527/jas.2014-8688

Uehara, T., and Park, J. T. (2008). Growth of *Escherichia coli*: significance of peptidoglycan degradation during elongation and septation. *J. Bacteriol.* 190, 3914–3922. doi: 10.1128/JB.00207-08

Uehara, T., Parzych, K. R., Dinh, T., and Bernhardt, T. G. (2010). Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *EMBO J.* 29, 1412–1422. doi: 10.1038/emboj.2010.36

United Nations–Department of Economic and Social Affairs/Population Division (2017). *World Population Prospects: the 2017 revision. Key findings and advance tables, ed.* United Nations New York: United Nations.

Van Soest, P. J. (1994). *Nutritional ecology of the ruminant*. 2nd. New York: Cornell University Press.

Vigors, S., O'Doherty, J. V., Kelly, A. K., O'Shea, C. J., and Sweeney, T. (2016). The Effect of Divergence in feed efficiency on the intestinal microbiota and the intestinal immune response in both unchallenged and lipopolysaccharide challenged ileal and colonic explants. *PLoS One* 11, e0148145. doi: 10.1371/journal.pone.0148145

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C. A., Hyslop, J. J., Ross, D. W., et al. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 16, 839. doi: 10.1186/s12864-015-2032-0

Warren, M. J., Raux, E., Schubert, H. L., and Escalante-Semerena, J. C. (2002). The biosynthesis of adenosylcobalamin (vitamin B12). *Nat. Prod. Rep.* 19, 390–412. doi: 10.1039/b108967f

Wang, A., Gu, Z., Heid, B., Akers, R. M., and Jiang, H. (2009). Identification and characterization of the bovine G protein-coupled receptor GPR41 and GPR43 genes. *J. Dairy Sci.* 92, 2696–2705. doi: 10.3168/jds.2009-2037

Zapun, A., Contreras-Martel, C., and Vernet, T. (2008). Penicillin-binding proteins and β-lactam resistance. *FEMS Microbiol. Rev.* 32, 361–385. doi: 10.1111/j.1574-6976.2007.00095.x

Zgurskaya, H. I., Krishnamoorthy, G., Ntreh, A., and Lu, S. (2011). Mechanism and function of the outer membrane channel TolC in multidrug resistance and physiology of enterobacteria. *Front. Microbiol.* 2, 1–13. doi: 10.3389/fmicb.2011.00189

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.3 Validation of previously identified biomarkers for feed conversion efficiency, appetite, and growth

After publication of the journal article in August of 2019, we performed a validation procedure, to evaluate the accuracy of the identification of the microbial genes identified as biomarkers for FCR, RFI, ADG, and DFI. This consisted of the random selection of groups of microbial genes containing the same number of microbial genes that had been identified previously as biomarker for each of the host traits, i.e., groups of 20, 17, 14, and 17 microbial genes were randomly selected from the original dataset and used in the evaluation of the accuracy of the identification of the previously identified biomarkers for FCR, RFI, ADG, and DFI, respectively. Using each set of randomly selected microbial genes, a partial least squares (PLS) model was fitted to predict the corresponding host performance trait. This procedure was repeated 100 times, including and excluding the fixed factors (diet and breed), and the results are presented in tables 1 and 2, respectively. The analyses were performed using the R software (v. 1.1.453) (R Core Team, 2021).

Table 1. *Percent variation of host performance traits feed conversion ratio, residual feed intake, average daily gain and daily feed intake explained by diet, breed, trial, and groups of randomly selected microorganisms*

Trait	Average of explained variation (%)	Standard deviation of explained variation (%)
FCR	6.15	2.33
RFI	4.25	1.55
ADG	4.60	1.74
DFI	4.45	1.40

FCR, RFI, ADG and DFI correspond to feed conversion ratio, residual feed intake, average daily gain, and daily feed intake, respectively. Average and standard deviation of explained variation (%) were obtained by averaging and calculating the standard deviation of the explained variation obtained from 100 partial least squares models for the prediction of each trait, based on the relative abundances of randomly selected groups of microbial genes, and including fixed factors (breed, diet, and trial).

PLS models including or excluding the fixed factors breed, diet, and trial, and groups of microbial genes randomly selected from the same dataset from which we had previously identified biomarkers of the host performance traits explained a substantially lower (on average less than 7%) proportion of the variation observed in the traits.

Table 2. *Percent variation of host performance traits feed conversion ratio, residual feed intake, average daily gain and daily feed intake explained by groups of randomly selected microorganisms*

Trait	Average of explained variation (%)	Standard deviation of explained variation (%)
FCR	6.67	2.94
RFI	5.14	3.10
ADG	4.41	1.88
DFI	5.17	2.80

FCR, RFI, ADG and DFI correspond to feed conversion ratio, residual feed intake, average daily gain, and daily feed intake, respectively. Average and standard deviation of explained variation (%) were obtained by averaging and calculating the standard deviation of the explained variation obtained from 100 partial least squares models for the prediction of each trait, based on the relative abundances of randomly selected groups of microbial genes.

These results further highlighted the close association between the biomarkers identified for the prediction of the host traits in our published article and each of the traits, which explained 63, 66, 73, and 65% of the variation of FCR, RFI, ADG and DFI, respectively.

3.4 Conclusions

There is a close association between the rumen microbiome and the host performance traits, particularly involving biochemical pathways such as those of cellulose and hemicellulose degradation, vitamin B12 synthesis, and amino acids metabolism, and microbial genes involved in these functions were relatively more abundant in highly efficient animals, whereas the microbiome of animals with lower feed efficiency showed higher relative abundances of microbial genes associated with LPS biosynthesis, cAMP resistance, and degradation of toxic compounds.

Focusing on the rumen microbial genes (i.e., the metagenome) provides very detailed information about which microbiome functions are altered between groups of hosts exhibiting different levels of a trait, or for example, subject to different treatments in an experimental setting. However, this study, as many others, was based on the microbiome compositions of samples collected at the abattoir, and knowledge is still lacking on whether rumen microbiome samples taken post-mortem are representative of previous moments of the animals' life. Therefore, whether the association between these microbial

genes and the bovine host traits is stable throughout the bovine's adult life is still unknown, which prompted the study presented in the next chapter of this thesis.

3.5 References

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.r-project.org/

Chapter 4 Temporal stability of the rumen microbiome and its associations with performance traits in beef cattle

4.1 Abstract

Studies focused on the rumen microbiome characterization and its association with host traits are very often based on one single microbiome sample, mostly collected at slaughter. However, whether this is representative of the rumen microbiome at earlier stages of growth is still unclear. In the present study, we investigated the temporal stability of the rumen microbiome, and its association with host performance traits, including feed conversion ratio (FCR), average daily weight gain (ADG), daily feed intake (DFI), residual feed intake (RFI), daily methane (CH_4) emissions (g/day) and CH_4 yield (g/kg of daily dry matter intake). A total of 20 beef cattle had their rumen digesta sampled at 6 timepoints, before inclusion of a diet additive (nitrate- or oil-based), at the start, mid, and end of a 56-day performance testing period (during which animals were tested for FCR, ADG, DFI, and RFI), after leaving the respiration chamber in which each steer was measured for its CH_4 emissions individually, and after slaughter. Microbial DNA was extracted from each sample, and whole metagenomic shotgun sequencing was performed to determine the abundance of microbial genera and genes. We compared microbiome profiles at both microbial genera and genes levels, throughout the sampling timepoints, evaluating alpha and beta diversity, correlation between timepoints, and using partial least squares (PLS) models to understand whether the associations between the rumen microbiome and host traits were stable throughout the finishing phase of cattle. The results showed overall that diversity indices of microbiome profiles were not affected by time, and Pearson correlations between microbial genera and genes vectorized matrices were, respectively, 71% and 81% between pre-additive and end test, and higher between other pairs of sampling timepoints. Microbiome compositions generated from samples taken at any sampling timepoint explained high percentages of the

variation of the performance traits, averaging between 55 ± 9% and 64 ± 10% (FCR and CH_4 emissions in g/day, respectively) when using microbial genera, and between 74 ± 7% and 86 ± 2% (ADG and CH_4 yield, respectively) when using microbial genes as predictors. The use of moderately to highly heritable microbial genes to predict the estimated breeding values of the performance traits revealed that the association between the microbiome and the host traits was strong over the entire finishing period, with PLS models explaining 67 ± 10% and 73 ± 8% (RFI and DFI, respectively). These longitudinal analyses revealed that the rumen microbiome remained remarkably stable, and that the associations between the rumen microbiome and the host performance traits are overall maintained throughout the finishing phase of beef cattle, indicating that microbiome data generated from samples taken even before the performance test could be used for breeding purposes.

4.2 Introduction

The increasing demand for foods rich in quality protein such as milk and meat emphasises the need for the improvement of ruminant production systems in terms of both economic efficiency and environmental impact. Ruminants have the unique ability to convert human undigestible feed, such as grass, into meat and milk, due to their symbiotic relationship with their rumen microbiome, composed of the rumen microbiota, i.e., all the microbes that inhabit the rumen, and the metagenome, i.e., their microbial genes. The rumen harbours bacteria, protozoa, and fungi, that ferment complex polysaccharides in the plant biomass, such as cellulose and hemicellulose, into volatile fatty acids (VFA), microbial protein, and vitamins, that the ruminant can absorb and/or further digest and utilize in its own development and growth. During the fermentation process, excess hydrogen can be produced that methanogenic archaea use to produce ATP, leading to methane (CH_4) emissions being released, mostly through eructation. Previous authors have shown the close association between the rumen microbiome and the host performance traits; Myer et al. (2015) reported increased relative abundances of Firmicutes in steers with greater growth rates, Roehe et al. (2016) identified 49 microbial genes that

explained 86% of the variation in feed conversion efficiency, Lima et al. (2019) found that sets of fewer than 20 rumen microbial genes explained more than 60% of the variation in feed efficiency traits including daily weight gain, appetite and feed conversion efficiency (chapter 3), and Auffret et al. (2020) identified microbial functional pathways such as biofilm formation, secretion system, and fucose sensing to be associated with higher feed conversion efficiency in beef cattle. Additionally, the association between the rumen microbiome and CH_4 emissions has been extensively explored; for example, Wallace et al. (2014) proposed that the Archaea:Bacteria ratio was a suitable biomarker for CH_4 emissions, Roehe et al. (2016) identified 20 microbial gene biomarkers that explained 81% of the variation in CH_4 emissions, Danielsson et al. (2017) observed increased relative abundances of *Methanobrevibacter ruminantium* and *Methanobrevibacter olleyae* in low CH_4 emitting dairy cattle, Auffret et al. (2018) showed that *Butyrivibrio* and *Pseudobutyrivibrio* were strongly associated with high CH_4 emissions in beef cattle, and Martínez-Álvaro et al. (2020) uncovered, based on complex interactions among rumen microbial communities and their genes, functional niches affecting CH_4 emissions, and that these emissions are mostly driven by microbial communities other than methanogenic archaea.

Most of the studies of associations between the microbiome and host traits are based on rumen samples taken at a single time point, usually immediately after slaughter. Previous studies on temporal stability of the rumen microbiome have found some fluctuations, e.g., Piao et al. (2014) showed that short-term shifts in the rumen microbiota profiles (specifically at 30min and 4h of incubation in the rumen) were associated to plant-biomass degradation rate, and Huws et al. (2016), based on 16S rRNA-based taxonomic compositions, showed that the colonization of fresh perennial ryegrass throughout an 8 hour period when feed particles enter the rumen of dairy cattle occurred fast, and that the microbiota profiles at 1 and 2h were significantly different from those at 4, 6, and 8 hours of particle incubation in the rumen. In a long-term study, Snelling et al. (2019) showed that the rumen microbiota was temporally stable

and suggested that, after a period of adaptation to potential dietary interventions, a single sample could be considered reasonably representative of the microbial communities. Additionally, the rumen microbiome is under the influence of the host genetics; Roehe et al. (2016) used a genetic model based on sire progeny groups and showed that information on microbiome profiles is crucial to identify and select animals with desirable characteristics for breeding programs, and Martínez-Alvaro et al. (2021) reported that rumen microbial genes and genera were significantly heritable, particularly those associated with CH_4 emissions, further confirming and specifying that a substantial portion of the host genetic control over its traits occurs through the control of its own microbiome profile.

In this study, we assessed the temporal stability of the rumen microbiome by focusing on the microbial communities and their microbial genes throughout the finishing period of 7 months of beef cattle. Furthermore, we investigated the stability of the prediction of host performance traits (feed conversion ratio, average daily weight gain, daily feed intake, and residual feed intake, FCR, ADG, DFI, and RFI, respectively) and CH_4 emissions traits (daily CH_4 emissions, in g/day, and yield, in g/kg daily dry matter intake) based on the rumen microbiota and metagenomic profiles collected at 6 sampling timepoints during growth.

4.3 Material and methods

This study was based on data collected from a previous animal trial investigating the effect of different diets and feed additives on animal performance and CH_4 emissions in different breeds of beef cattle during the finishing phase.

4.3.1 Experimental design, animals, and diets

The animal trial was a 2 × 2 × 3 factorial design performed that included 84 animals from crossbred Charolais (Chx, n=42) and purebred Luing (Lu, n=42) breeds. These steers were allocated to one of two basal diets (consisting of forage:concentrate ratios of 520:480, and 84:916 g/kg dry matter basis; forage,

and concentrate, respectively). On the allocation day, these animals weighted on average 414.1 ± 35.7 kg. Before adaptation to the basal diet, all animals were fed the forage diet. Half of the steers were then adapted to the concentrate diet in a stepwise manner over a period of 4 weeks. Animals fed each basal diet were then allocated to one of three treatments; the control (CTL), the supplementation of 21.5g nitrate/kg dry matter (NIT), and the use of high oil content rapeseed cake (RSC). Animals were adapted to the NIT and RSC treatments in a stepwise manner over a period of four weeks. During the trial, the steers were offered diets ad libitum, at approximately 1.05 times the average daily feed intake. Diet composition and more details on the feeding trial can be found in Duthie et al. (2016).

The experiment was conducted at Scotland's Rural College (SRUC) Beef and Sheep Research Centre in Edinburgh in 2013. The experimental protocol was approved by SRUC's Animal Welfare and Ethical Review Body, the Animal Experiments Committee, and was conducted in accordance with the requirements of the United Kingdom Animals (Scientific Procedures) Act, 1986.

4.3.2 Animals selected for rumen digesta sampling

To limit whole metagenomic sequencing cost for the longitudinal samples, a total of 20 animals were selected for rumen digesta sampling (10 Chx, and 10 Lu). Six animals from each breed were fed concentrate, which were balanced for the three treatment groups (CTL, NIT, or RSC), whereas 4 animals from each breed were fed forage, and balanced for 2 treatment groups (CTL or NIT).

4.3.3 Bovine host performance traits

Performance traits were measured during a 56-day testing period. DFI was assessed by measuring dry matter intake (DMI, kg/day), which was recorded using electronic feeding equipment (Insentec, Marknesse, The Netherlands). Body weight (BW) was measured weekly using a calibrated weight scale (before fresh feed was offered). ADG was modelled by linear regression of BW against test date. FCR was calculated as average DMI (kg/day) divided by

ADG. Daily CH_4 emissions were measured in respiration chambers, in which each steer remained for 3 days, whereas the final 48 hours were used in the estimation of daily CH_4 emissions. Daily CH_4 emissions was divided by the average DMI measured within the respiration chambers, resulting in CH_4 yield. A more detailed description can be found in Duthie et al. (2016).

4.3.4 Rumen digesta sampling timepoints

All animals had their rumen digesta sampled after the adaptation to the basal diets but before introduction of feed additives (pre-additive) at an average of 418 ± 32 days old, at the start (average of 460 ± 32 days old), mid-point (average of 492 ± 32 days old), and end (average 520 ± 32 days old) of the 56-day performance test period (start, mid, and end RFI, respectively), on the day the animal left the respiration chambers (chamber), and at the abattoir within 2 hours of the animal slaughter (slaughter; Figure 1).

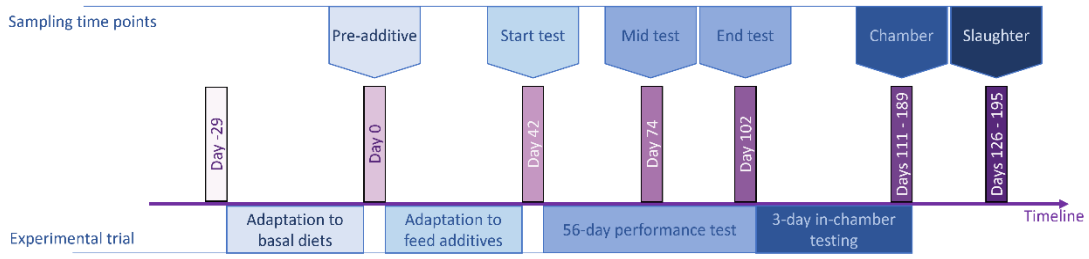


Figure 1. *Schematic representation of the experimental timeline. Test refers to the 56-day performance testing period.*

At each sampling, approximately 50 mL of rumen liquid were taken by inserting a stomach tube (16 × 2700 mm Equivet Stomach Tube, Jørgen Kruuse A/S, Langeskov, Denmark) nasally and aspirating manually. This liquid was filtered through two layers of muslin and 5 mL strained rumen fluid were mixed with 10 ml phosphate buffered saline containing glycerol (30% v/v). These samples were stored at −20°C between collection and analysis. Animals were slaughtered in a commercial abattoir where two samples of rumen digesta (~50 ml) were collected immediately after the rumen was opened to be drained. The

slaughterhouse sample collection process results in well-mixed samples of rumen contents.

4.3.5 Whole metagenomic sequencing

DNA was extracted from the samples of rumen digesta obtained from 20 animals following the methodology described in Rooke et al. (2014). Illumina TruSeq libraries were prepared from genomic DNA and sequenced on Illumina HiSeq systems 4000 (8 animals) or Illumina PE150 (12 animals) by Edinburgh Genomics (Edinburgh, UK). Paired-end reads (2 × 150 bp) were generated, resulting in between 16 and 42 GB per sample (between 55 and 140 million paired reads). To measure the abundance of known functional MGs whole metagenome sequencing reads were quality trimmed using Fastp (Chen et al., 2018) and assembled using MEGAHIT (Li et al., 2015). Proteins were predicted using Prodigal (Hyatt et al., 2010) and searched against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/ko.html) (version 2020-10-04) (Kanehisa and Goto, 2000) using KofamScan database (Aramaki et al., 2020). Hits that passed KofamScan's default thresholds were assigned to KEGG orthologous groups (KO). Proteins that passed the threshold for multiple KOs were grouped separately, as were those that did not have a hit. The resulting KO grouping corresponded to a highly similar group of sequences. For phylogenetic annotation of rumen samples, we followed the same pipeline as described in Martínez-Álvaro et al. (2021a). Briefly, the sequence reads of the samples were aligned to a database including cultured genomes from the Hungate 1000 collection (Seshadri et al., 2018) and Refseq genomes (Pruitt et al., 2007) using Kraken software (Wood and Salzberg, 2014).

4.3.6 Statistical analyses

4.3.6.1 Data cleaning and transformation

The microbiome data included a total of 1178 microbial genera 6916 microbial genes. Microbial genera and genes were removed from the dataset if they were absent from at least one animal and/or if they had average relative

abundance lower than 0.001%, leaving a total of 1050 microbial genera and 1902 microbial genes for further analyses.

To deal with the compositional nature of microbiome data, a transformation was applied using the additive logratio (ALR; Gloor et al., 2017), with denominators *Oribacterium* and *K00858* for microbial genera and genes, respectively. The denominators were selected following the criteria proposed in Greenacre et al. (2021), and previously applied in Martínez-Álvaro et al. (2021a, 2021c), based on high Procrustes correlation between the ALR-transformed data and the centred logratio-transformed data (ensuring isometry between samples), and the lowest variance in the denominator (simplifying the interpretation of results).

4.3.6.2 Diversity indices

To investigate the temporal stability of the samples in terms of diversity within sample, and dissimilarity from other samples, we used the original datasets (in absolute counts) including 1178 microbial genera and 6916 microbial genes to calculate the adjusted Shannon index of each sample in each sampling timepoint and the Bray-Curtis dissimilarity between all samples at all timepoints, respectively, using the vegdist() function included in the vegan package (Oksanen et al., 2019). We compared Shannon indices using linear mixed models with sampling timepoints as fixed effect and animal ID as random effect (independently distributed with mean 0 and individual variance multiplied by an identity matrix of 20 × 20 order), using the lme4 package (Bates et al., 2015) in R version 1.4.1103 (R Core Team, 2021). Post hoc tests to compare Shannon indices from sampling timepoints in a pairwise manner were performed using the package emmeans (Russell et al., 2021), in R. Bray-Curtis dissimilarities in different timepoints were compared in a PERMANOVA calculated using the adonis2() function from the vegan package in R, and a post hoc pairwise comparison of the dissimilarities in each sampling timepoint was performed using the pairwise.adonis2() function from the pairwiseAdonis package (Martinez Arbizu, 2020) in R. Because the PERMANOVA can result in significant differences due to either significant differences in the centroids or

due to heteroscedasticity of the groups of samples, we calculated the homogeneity of group dispersions (i.e. dispersion within sampling timepoints) using the betadisper() function of the vegan package and compared them in an anova in R. Additionally, Bray-Curtis dissimilarities were projected using the nonmetric multidimensional scaling (NMDS), with random seeds set at 10403 and 432 (first and second elements of the .Random.seed vector in R) for microbial genera and genes level analyses, respectively.

4.3.6.3 Temporal stability of microbial genera and genes throughout the finishing period

To evaluate the temporal stability of the abundances of microbial genera and genes throughout the sampling timepoints, we fitted 1049 and 1901 linear mixed models using as dependent variables the ALR-transformed abundances (ALR-As) of microbial genera and genes, respectively, timepoint as fixed effect, and animal ID as random effect (independently distributed with mean 0 and individual variance multiplied by an identity matrix of 20 × 20 order). P-values testing for significance of different ALR-As between timepoints were corrected by the Bonferroni method, and significance was assessed at p-value < 0.05. These analyses were done using the lme4 (Bates et al., 2015) and car (Fox and Weisberg, 2019) packages in R.

Additionally, we compared the microbial genera and gene datasets by calculating the Pearson correlation between the vectorized microbial genus and microbial gene ALR-As matrices generated from the samples collected at each sampling timepoint in a pairwise manner. These analyses were performed in R.

4.3.6.4 Prediction ability of performance and methane emissions traits based on data generated from different sampling timepoints

Partial Least Squares (PLS) models were used to understand which of the 1049 microbial genera and 1901 microbial genes were most associated to each trait in each timepoint. The traits included in the models were the residuals of FCR, ADG, DFI, RFI, daily CH_4 emissions, and CH_4 yield, obtained from linear models regressing each trait on the animals' weights as measured

on the day they were allocated to the basal diet. These analyses were performed using the mixOmics (Cao et al., 2020) package in R. Variable importance in projection (VIP) scores were obtained for each explanatory variable in each predictive model using 2 components. Regression coefficients from the first component were used to describe the direct relationship between each predictor variable (microbial genera or genes) and the trait. Significance levels of the regression coefficients were obtained using the Jack-Knife procedure, based on intervals of average ± 2 standard deviations (regression coefficients distribution obtained from leave-one-out cross validation).

Additionally, we investigated whether microbial genera and genes measured at slaughter previously identified as highly associated (VIP ≥ 1) to performance and methane emissions traits by our group would still be identified as important using the datasets generated from samples at earlier sampling timepoints. For this analyses we included the microbial genera and genes identified for prediction of methane yield by Martínez-Álvaro et al. (2020), the microbial genes identified for prediction of FCR, ADG, DFI, and RFI by Lima et al. (2019) and the microbial genes identified for prediction of FCR and methane yield by Roehe et al. (2016) (these manuscripts are identified as MA, L, and R, respectively). In these studies, the authors identify host performance traits' biomarkers based on the investigation of the relative abundance of microbial genera and genes, using PLS and/or correlation networks, with enrichment analyses.

4.3.6.5 Stability of the association of host-genomically influenced microbial genes with the estimated breeding values of host performance traits

We repeated the analyses exclusively based on the host-genomically influenced functional core microbiome composed of 443 microbial gene ALR-As defined in Martínez-Álvaro et al. (2021b) identified in the datasets generated from all sampling timepoints of our analyses. These microbial genes presented significant host-genomic effects after a stringent multitest correction, estimated by using a larger data set (n=359 animals) with metagenomics and host genomic information (described in Martínez-Álvaro et

al. (2021b)) which included the animals in this study. Additionally, estimated breeding values (EBVs) were calculated based on records in the complete database (n=359 animals), using univariate GBLUP analysis resolved with BGLR software (Pérez and De Los Campos, 2014). The univariate models included correction for systematic effects (diet, breed, and trial) and followed the same assumptions as in Martínez-Álvaro et al. (2021b). A total of 443 microbial gene ALR-As previously estimated to have moderate to high heritability generated from samples collected from each sampling timepoint were evaluated in the present study for their stable association with the EBVs of the performance traits FCR, ADG, DFI, RFI, and CH_4 yield. These analyses were performed by using the ALR-As as predictor variables in PLS models that included the EBVs of the performance traits as dependent variable.

4.4 Results

4.4.1 Diversity indices

The diversity indices were remarkably stable throughout the finishing period. No significant differences in diversity indices were observed in the analyses at microbial genera level. At the microbial genes level, only those from slaughter samples had significantly higher Shannon diversity than mid test samples (averages of 0.891 and 0.880, respectively, p-value = 0.02), whereas no significant differences in diversity indices were observed for any other pairwise comparison (Figure 2).

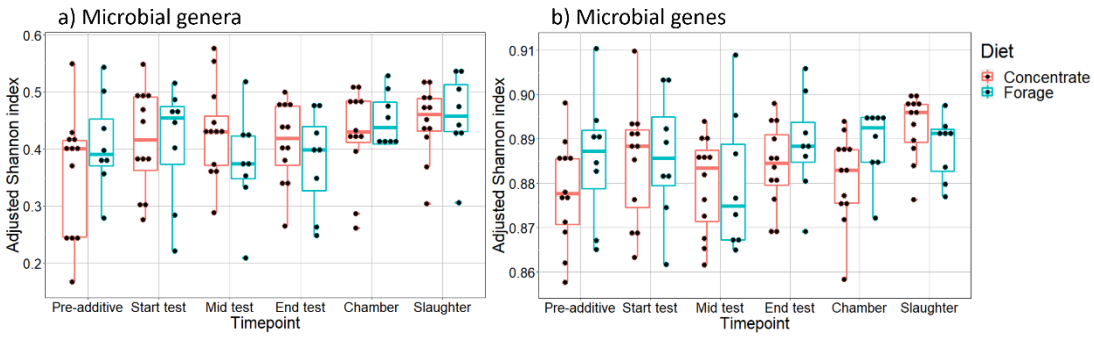


Figure 2. *Adjusted Shannon index for a) microbial genera and b) microbial genes throughout the sampling timepoints.*

The PERMANOVA analysis of Bray-Curtis dissimilarities revealed significant differences between the dissimilarity matrices calculated based on abundances of microbial genera at different sampling timepoints, due to differences between the centroids of those dissimilarity matrices at different timepoints (dispersions from the centroids based on Euclidian distance within each timepoint were not significantly different between timepoints). The pairwise analyses showed the differences occurred mostly between dissimilarity matrices based on slaughter data and those obtained from other timepoints, and that, overall, timepoint had little effect on the centroids of the BC dissimilarity matrices, explaining only 0.143 or lower of the total variance in the PERMANOVA models (Table 1).

Table 1. *Comparison of distance matrices between different sampling timepoints based on Bray-Curtis dissimilarity indices.*

Timepoint dataset	Microbial genera		Microbial genes	
PERMANOVA	Explained variance	P-value	Explained variance	P-value
All datasets	0.085	0.001	0.042	0.419
Heteroscedasticity	n.a.	0.378	n.a.	0.060
PERMANOVA Pairwise comparisons	Explained variance	P-value	Explained variance	P-value
Pre-additive vs Start test	0.039	0.178	0.014	0.614
Pre-additive vs Mid test	0.028	0.327	0.019	0.416
Pre-additive vs End test	0.037	0.188	0.023	0.393
Pre-additive vs Chamber	0.049	0.086	0.011	0.713
Pre-additive vs Slaughter	0.143	0.002	0.084	0.025
Start test vs Mid test	0.006	0.962	0.005	0.941
Start test vs End test	0.006	0.972	0.006	0.895
Start test vs Chamber	0.016	0.699	0.006	0.876
Start test vs Slaughter	0.101	0.003	0.058	0.084
Mid test vs End test	0.007	0.964	0.004	0.970
Mid test vs Chamber	0.011	0.832	0.010	0.687
Mid test vs Slaughter	0.114	0.001	0.062	0.075
End test vs Chamber	0.017	0.689	0.009	0.722
End test vs Slaughter	0.106	0.001	0.054	0.103
Chamber vs Slaughter	0.097	0.002	0.057	0.086

n.a. refers to non-applicable; shaded cells represent significant differences.

Additionally, the NMDS plots of Bray-Curtis dissimilarities based on microbial genera and genes showed no distinctive clustering of samples by sampling timepoint (Figure 3).

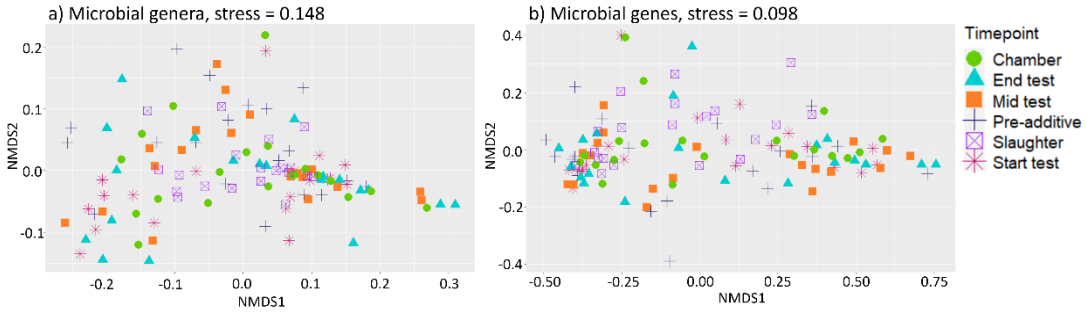


Figure 3. *Non-metric multi-dimensional scaling (NMDS) plot of Bray-Curtis distances between samples at a) microbial genera and b) microbial genes level.*

4.4.2 Temporal stability of microbial genera and genes

The linear mixed models regressing the ALR-As of each microbial genera and gene on sampling timepoint, using animal ID as random effect revealed that 11 and 26 microbial genera and genes, respectively, had significantly different abundances over timepoints; meaning that the abundances of 99% of the microbial genera and of the microbial genes were stable throughout the 6 timepoints. The 5 most abundant microbial genera and genes did not show significant differences between timepoints (Figure 4).

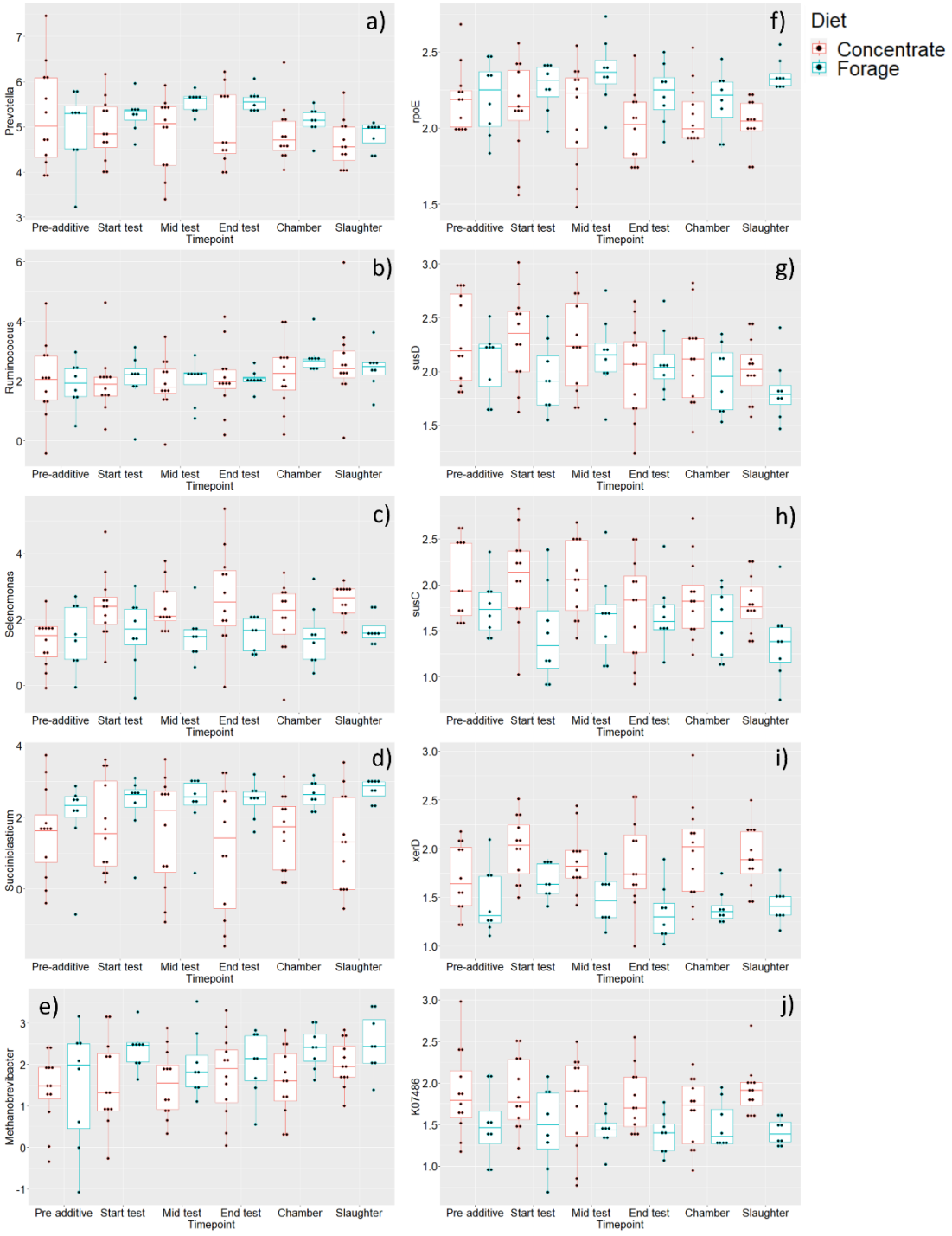


Figure 4. *ALR-transformed abundances of most abundant microbial genera a) Prevotella, b) Ruminococcus, c) Selenomonas, d) Succiniclasticum, and e) Methanobrevibacter, and microbial genes f) DNA-directed RNA polymerase subunit delta (rpoE), g) starch-binding outer membrane protein, SusD/RagB family (susD), h) TonB-dependent starch-binding outer membrane protein SusC (susC), i) integrase/recombinase XerD (xerD), and j) transposase (K07486), throughout all sampling timepoints*

The Pearson correlations between the vectorized microbiome data showed that those generated from samples taken after the animals left the respiration

chambers (microbial genera) and at mid test (microbial genes) were the ones that mostly resembled the microbiome generated from slaughter samples (Table 2). However, all pairwise analyses resulted in strong correlations, with the lowest being 71% between pre-additive and end test for microbial genera and 81% between pre-additive and end test for microbial genes data.

Table 2. *Correlations (%) between vectorized matrices of ALR-transformed abundances of microbial genera and genes generated from each sampling timepoint*

Timepoints	Pre-additive	Start test	Mid test	End test	Chamber	Slaughter
Pre-additive		80.69	72.04	70.58	80.75	78.05
Start test	81.85		77.10	79.14	82.46	81.19
Mid test	81.52	84.25		87.15	81.69	82.56
End test	81.36	83.35	89.20		82.72	80.75
Chamber	82.59	85.27	83.11	84.57		85.10
Slaughter	83.55	85.21	85.82	84.98	84.57	

Pairwise Pearson correlations between microbial genera at different sampling timepoints are above the diagonal, while those based on microbial gene at different sampling timepoints are below the diagonal. Shaded cells represent the sampling timepoints with highest correlation to slaughter samples.

4.4.3 Prediction ability of microbial genera and genes throughout the finishing phase

The PLS models predicting the performance traits based on the microbiome at each different sampling timepoint revealed that microbial genera or microbial genes explained very similar amounts of variance of the phenotypes throughout the different timepoints (Table 3), with averages of 55.4 ± 9.1%, 56.3 ± 9.2%, 60.9 ± 12.2%, 59.4 ± 12.3%, 60.2 ± 4.6% and 64.3 ± 10.3% (microbial genera) and 74.1 ± 6.3%, 73.8 ± 7.3%, 80.3 ± 4.2%, 74.2 ± 6.2%, 85.7 ± 2.2% and 82.5 ± 3.3% (microbial genes) for FCR, ADG, DFI, RFI, CH_4 yield and daily CH_4 emissions, respectively.

Table 3. *Proportion of explained variance of performance and methane emissions traits predicted by microbial genera or genes generated from rumen samples taken at each sampling timepoint*

	Trait	FCR		ADG		DFI		RFI		CH_4 yield		Daily CH_4 emissions	
	Timepoint	X (%)	Y (%)	X (%)	Y (%)	X (%)	Y (%)	X (%)	Y (%)	X (%)	Y (%)	X (%)	Y (%)
Microbial genera	Pre-additive	64.71	44.77	63.57	54.33	62.82	71.46	62.37	66.59	65.88	54.10	62.66	76.99
	Start test	57.47	70.19	57.69	67.77	54.02	78.33	65.11	44.74	69.51	57.44	68.10	51.42
	Mid test	54.20	56.21	31.71	62.64	67.20	61.41	24.68	80.17	66.72	58.46	65.18	74.16
	End test	61.63	52.41	53.88	61.24	62.35	53.80	64.06	56.77	62.80	60.43	65.30	63.93
	Chamber	57.51	60.35	60.91	47.11	65.18	45.26	58.10	54.55	67.12	66.59	66.15	53.93
	Slaughter	57.34	48.65	59.94	44.51	52.64	55.07	58.95	53.70	64.86	64.37	52.50	65.52
Microbial genes	Pre-additive	33.45	82.90	34.08	82.59	24.53	88.10	36.41	83.29	23.36	88.01	35.05	80.57
	Start test	39.61	65.19	38.35	63.78	33.17	77.45	34.90	64.41	33.99	85.59	33.61	83.41
	Mid test	37.54	70.95	39.42	73.31	42.54	78.25	42.88	72.21	40.28	87.15	41.06	88.47
	End test	39.97	78.47	40.78	78.79	43.77	79.19	41.26	73.15	37.22	84.33	44.30	82.95
	Chamber	30.12	76.23	30.82	77.72	28.55	82.13	31.77	77.00	33.70	86.99	32.27	79.35
	Slaughter	34.60	70.89	38.24	66.87	34.81	76.84	35.07	75.00	40.66	81.99	37.17	80.12

FCR, ADG, DFI, RFI, CH_4 yield, and daily CH_4 emissions refer to feed conversion ratio, average daily weight gain, daily feed intake, residual feed intake, methane yield (g/Kg DMI) and daily methane emissions (g/day) respectively. X (%) and Y (%) represent the percentage of explained variation in the microbiome and the trait, respectively, based on Partial Least Squares models using 2 components and including a total of 1049 microbial genera ALR-transformed abundances (upper table part) and 1901 microbial gene ALR-transformed abundances (lower table part). The shaded cells represent the timepoint that explained the highest variance of each trait.

The analyses of the most important microbial genera and genes revealed that a high proportion of microbial genera and genes identified as important (VIP ≥ 1) for prediction of traits based on slaughter samples, would also be important for the prediction of those traits using microbiome data generated from other timepoints (Table 4).

Table 4. *Number of microbial genera or genes important to predict each host performance and methane emissions trait at each timepoint, and percentage of those genera and genes at each given timepoint also important in slaughter samples*

	Trait	FCR		ADG		DFI		RFI		CH_4 yield		Daily CH_4 emissions	
	Timepoint	VIP≥1	SL(%)	VIP≥1	SL(%)	VIP≥1	SL(%)	VIP≥1	SL(%)	VIP≥1	SL(%)	VIP≥1	SL(%)
Microbial genera	Pre-additive	355	44.4	356	42.3	349	36.4	328	34.1	352	32.0	358	26.2
	Start test	362	34.2	384	40.5	349	26.6	409	37.5	384	36.7	438	65.6
	Mid test	448	44.6	427	42.6	454	50.7	418	35.9	405	58.3	431	68.8
	End test	381	38.6	334	29.7	425	40.0	404	36.3	443	54.1	443	65.6
	Chamber	448	46.2	433	47.1	390	35.9	418	35.0	390	61.9	402	50.1
	Slaughter	383	100.0	333	100.0	418	100.0	320	100.0	412	100.0	381	100.0
Microbial genes	Pre-additive	723	43.3	651	37.8	584	29.7	817	51.5	714	26.8	666	25.4
	Start test	819	46.6	807	45.5	769	43.2	789	41.9	861	54.5	820	57.6
	Mid test	650	29.6	717	43.1	901	56.3	877	52.1	868	55.3	910	64.0
	End test	805	51.0	860	54.5	926	61.8	908	61.4	879	60.3	849	61.9
	Chamber	801	42.9	834	45.5	770	43.2	829	52.9	806	59.0	858	56.9
	Slaughter	802	100.0	792	100.0	704	100.0	823	100.0	851	100.0	861	100.0

FCR, ADG, DFI, RFI, CH_4 yield, and daily CH_4 emissions refer to feed conversion ratio, average daily weight gain, daily feed intake, residual feed intake, methane yield (g/Kg DMI) and daily methane emissions (g/day) respectively. VIP ≥ 1 represents the number of microbial genera and genes with variable importance in projection (VIP) score above 1. SL (%) was calculated as the percentage of microbial genera and genes important for prediction of the trait (VIP ≥ 1) obtained at each timepoint that were also important for prediction of those traits at slaughter. Shaded cells represent the highest SL (%) for each trait. VIPs were obtained from Partial Least Squares prediction models using 2 components and including a total of 1049 microbial genera ALR-transformed abundances (upper table part) and 1901 microbial genes ALR-transformed abundances (lower table part) as explanatory variables.

The consistency of the relationship between the microbial genera and genes that explained the variance observed in performance and CH_4 emissions traits throughout all timepoints was assessed by analysing the significant regression coefficients (based on Jack-Knife intervals) of the microbial genera and genes important (VIP ≥ 1) for the prediction in each PLS model (Table 5). The results showed that the direction of the association (regression coefficient positive or negative) between the microbial genera and genes and the trait is mostly maintained, suggesting that the relationship between the explanatory variables and the host performance traits is mostly stable throughout time. Some changes in the direction of association could be due to the many correlations among the microbial genera or genes.

Table 5. *Regression coefficients of microbial genera and genes important for the prediction of each performance and methane emissions traits in all timepoints*

Trait	Microbial genus/ gene	Pre-additive	Start test	Mid test	End test	Chamber	Slaughter
FCR	*Candidatus Hamiltonella*	1.12E-03	-2.99E-04	-7.69E-04	-1.54E-03	-2.16E-03	-6.14E-04
	Algibacter	3.85E-05	-5.47E-04	4.91E-03	7.27E-04	1.54E-03	2.09E-03
	rbsK (K00852)	2.16E-03	1.20E-03	3.15E-03	-1.23E-04	-3.25E-05	-1.40E-04
	cpaB (K02279)	-9.55E-04	9.72E-04	-3.26E-04	1.35E-03	-5.92E-04	1.62E-03
	ltrA (K00986)	3.50E-03	1.52E-03	-1.25E-03	1.62E-03	-9.83E-04	1.84E-03
	ABC.PE.S (K02035)	1.98E-03	1.36E-03	-5.22E-04	-3.79E-04	9.88E-05	1.42E-03
ADG	*Dehalococcoides*	-9.81E-04	-8.00E-04	-2.09E-03	-9.90E-04	-2.34E-03	-1.80E-03
	Dichelobacter	2.73E-04	2.08E-04	-1.16E-03	-7.19E-06	-1.45E-03	-1.68E-03
	ABC.PE.S (K02035)	-2.83E-03	-1.35E-03	-3.68E-04	-4.62E-04	-2.85E-04	-1.03E-03
	glmM (K03431)	-1.90E-03	-9.42E-04	-8.73E-04	-8.99E-04	-4.51E-04	-1.07E-03
	ltrA (K00986)	-3.88E-03	-1.63E-03	-1.01E-04	-1.03E-03	7.38E-04	-1.44E-03
	ATPF1G (K02115)	-2.14E-03	-1.04E-03	-3.22E-04	-1.88E-04	-1.62E-04	2.02E-04
DFI	*Eubacterium*	-1.23E-03	-3.04E-04	-1.77E-03	-9.72E-04	-2.24E-03	-7.81E-03
	Methanobrevibacter	-2.42E-04	1.11E-03	2.33E-03	1.27E-03	2.73E-03	2.97E-03
	Cryptococcus	-2.44E-04	8.33E-05	2.63E-03	1.54E-03	2.03E-03	2.69E-03
	Scheffersomyces	1.68E-04	2.33E-04	2.00E-03	1.47E-03	2.13E-03	2.65E-03
	virD4 (K03205)	2.18E-03	-1.45E-03	-1.12E-03	-1.27E-03	-1.25E-03	-1.18E-03
	spoVK (K06413)	-3.95E-03	-1.14E-04	9.42E-04	9.90E-04	1.13E-03	2.15E-04
	paaK (K01912)	-2.42E-03	-1.63E-04	1.09E-03	8.53E-04	5.44E-04	1.34E-04
	iorB (K00180)	-2.19E-03	3.23E-04	1.09E-03	8.00E-04	1.12E-03	2.56E-04
	higB (K19166)	5.90E-03	2.00E-03	8.56E-04	1.13E-03	1.17E-03	1.63E-03
	ehaD (K14095)	2.09E-03	1.43E-03	1.17E-03	8.50E-04	1.31E-03	4.21E-04
RFI	*Azorhizobium*	-4.02E-04	-1.06E-03	-2.42E-03	-2.64E-03	-2.39E-03	-3.86E-05
	Roseburia	-3.11E-03	4.05E-04	-2.91E-03	-2.59E-03	-2.58E-03	-6.06E-04
	Jonquetella	5.04E-05	-1.09E-03	-3.10E-03	-2.99E-03	-2.57E-03	-2.11E-04
	pflX (K04070)	-1.06E-03	9.26E-04	-4.75E-04	-9.21E-04	-1.53E-03	-3.44E-04
	metA (K00651)	-1.21E-03	4.31E-04	-9.57E-04	-9.69E-04	-1.26E-03	-1.68E-03
	wecC (K02472)	-1.36E-03	4.83E-04	9.14E-04	3.91E-04	-4.56E-05	-1.48E-04
	ftnA (K02217)	8.85E-05	-1.44E-04	-8.27E-04	-1.64E-04	3.87E-04	1.73E-04
	EGD2 (K03626)	1.11E-03	1.29E-03	1.03E-03	8.98E-04	1.60E-03	1.13E-03
	RP-L30e (K02908)	9.73E-04	1.30E-03	1.04E-03	9.23E-04	1.80E-03	1.07E-03
Daily CH_4 emissions	*Eubacterium*	-1.26E-03	-1.93E-03	-3.40E-03	-1.74E-03	-1.23E-03	-4.27E-03
	Candidatus Methanoperedens	-1.98E-03	-2.48E-03	-2.61E-03	-1.97E-03	-1.13E-03	-4.51E-03
	Mitsuokella	-1.20E-03	-2.23E-03	-3.40E-03	-2.69E-03	-1.70E-03	-4.34E-03
	Cryptobacterium	-2.74E-04	-7.82E-04	-2.59E-03	-1.83E-03	-1.21E-03	-2.54E-03
	Dickeya	-1.39E-03	-1.81E-03	-2.71E-03	-1.62E-03	-7.17E-04	-3.51E-03
	Acidaminococcus	-1.42E-03	-1.96E-03	-3.06E-03	-2.04E-03	-1.36E-03	-3.17E-03
	Faecalitalea	-2.33E-03	-2.51E-03	-2.71E-03	-1.48E-03	-6.80E-04	-3.19E-03
	Methanoregula	3.97E-04	-2.15E-03	-2.87E-03	-2.21E-03	-5.95E-04	-3.79E-03
	Zygosaccharomyces	-5.07E-05	1.22E-03	2.58E-03	1.76E-03	1.54E-03	2.42E-03
	Blastomyces	-1.18E-05	1.43E-03	2.94E-03	1.99E-03	1.65E-03	2.29E-03
	Plasmodium	-1.14E-04	1.44E-03	2.86E-03	1.84E-03	1.77E-03	2.21E-03

	Neurospora	-1.74E-04	1.34E-03	2.38E-03	2.13E-03	1.67E-03	2.23E-03
	Lodderomyces	-9.33E-05	1.32E-03	2.38E-03	1.75E-03	1.79E-03	2.09E-03
	Saccharomyces	-8.08E-05	1.08E-03	2.78E-03	1.63E-03	1.84E-03	2.13E-03
	Halapricum	2.23E-03	2.27E-03	3.45E-03	2.44E-03	2.14E-03	4.03E-03
	Sarcina	4.15E-04	1.39E-03	2.46E-03	2.14E-03	2.01E-03	3.34E-03
	Calothrix	1.63E-04	1.58E-03	3.08E-03	1.89E-03	2.13E-03	3.01E-03
	Nodularia	3.51E-04	1.48E-03	2.94E-03	1.81E-03	2.12E-03	2.89E-03
	Haloquadratum	3.35E-04	1.34E-03	3.27E-03	2.08E-03	1.42E-03	2.79E-03
	Isaria	3.41E-04	1.64E-03	2.53E-03	1.89E-03	2.26E-03	2.68E-03
	Anabaena	4.37E-04	1.49E-03	3.17E-03	2.22E-03	1.70E-03	2.47E-03
	Histoplasma	9.12E-05	1.53E-03	2.84E-03	1.91E-03	1.75E-03	2.37E-03
	Coccidioides	2.20E-05	1.56E-03	2.51E-03	2.10E-03	1.87E-03	2.19E-03
	Emticicia	9.27E-05	1.65E-03	2.44E-03	1.50E-03	1.40E-03	2.17E-03
	Trypanosoma	1.07E-04	1.43E-03	2.72E-03	1.94E-03	1.81E-03	2.11E-03
	Microsporum	2.33E-04	1.69E-03	2.65E-03	2.06E-03	1.70E-03	2.01E-03
	Marssonina	7.90E-05	1.20E-03	2.88E-03	2.00E-03	1.54E-03	2.02E-03
	citF (K01643)	-2.17E-03	-1.21E-03	-9.74E-04	-9.09E-04	-1.53E-03	-1.15E-03
	sucD (K01902)	-1.82E-03	-1.07E-03	-1.07E-03	-1.12E-03	-1.24E-03	-1.04E-03
	fliJ (K02413)	2.31E-03	-1.22E-03	-1.06E-03	-1.32E-03	-1.21E-03	-1.46E-03
	dinJ (K07473)	1.95E-03	-1.14E-03	-1.26E-03	-1.39E-03	-3.02E-04	-1.21E-03
	gidA (K03495)	2.03E-03	-1.10E-03	-8.82E-04	-9.76E-04	-1.19E-03	-1.29E-03
	fliQ (K02420)	1.98E-03	-1.09E-03	-9.85E-04	-1.32E-03	-1.07E-03	-1.33E-03
	pncC (K03742)	3.94E-03	-1.46E-03	-7.92E-04	-1.35E-03	-1.53E-03	-1.32E-03
	bioD (K01935)	1.96E-03	-1.39E-03	-1.19E-03	-1.27E-03	-1.28E-03	-1.16E-03
	yfbR (K08722)	2.05E-03	-4.70E-04	-1.31E-03	-1.34E-03	2.50E-04	-9.50E-04
	fic (K04095)	1.75E-03	1.71E-04	-3.89E-04	-1.33E-03	4.82E-04	-1.16E-03
	hemD (K01719)	-1.71E-03	1.20E-03	1.16E-03	1.34E-03	1.22E-03	1.33E-03
	menI (K19222)	-3.91E-03	1.24E-03	8.99E-04	1.36E-03	1.18E-03	9.29E-04
	K09922	8.68E-05	4.25E-04	9.88E-04	9.28E-04	1.38E-03	1.14E-03
CH_4 yield	*Candidatus Phytoplasma*	9.48E-04	9.82E-04	1.44E-03	1.18E-03	1.79E-03	1.65E-03
	Marssonina	6.13E-04	1.50E-03	2.40E-03	1.60E-03	1.90E-03	1.83E-03
	Turneriella	1.40E-03	1.07E-03	1.46E-03	5.60E-05	1.38E-03	1.88E-03
	Nakaseomyces	1.28E-03	1.38E-03	2.24E-03	1.29E-03	1.85E-03	2.02E-03
	Blastomyces	8.13E-04	1.58E-03	2.26E-03	1.61E-03	2.02E-03	2.02E-03
	Trypanosoma	1.00E-03	1.41E-03	2.11E-03	1.54E-03	2.04E-03	1.96E-03
	Rhodotorula	8.93E-04	8.03E-04	1.34E-03	1.43E-03	1.35E-03	1.85E-03
	Zygosaccharomyces	8.27E-04	1.47E-03	2.08E-03	1.39E-03	1.90E-03	2.14E-03
	Pichia	1.02E-03	1.72E-03	1.88E-03	1.29E-03	1.93E-03	2.00E-03
	Saprolegnia	3.32E-04	1.85E-03	1.50E-03	1.52E-03	1.53E-03	2.33E-03
	Candidatus Protochlamydia	9.72E-04	1.49E-03	1.78E-03	1.45E-03	1.63E-03	1.58E-03
	arsC (K03741)	-1.80E-03	-1.34E-03	-1.08E-03	-9.26E-04	-1.60E-03	-1.34E-03
	ppdK (K01006)	-1.49E-03	-8.30E-04	-9.69E-04	-9.70E-04	-1.57E-03	-1.19E-03
	K07118	-2.50E-03	-1.48E-03	-1.03E-03	-1.07E-03	-1.83E-03	-1.42E-03

FCR, ADG, DFI, RFI, CH_4 yield, and daily CH_4 emissions refer to feed conversion ratio, average daily weight gain, daily feed intake, residual feed intake, methane yield (g/Kg DMI) and daily methane emissions (g/day) respectively. Regression coefficients extracted from the first component of 2 component Partial Least Squares prediction models based on 1049 microbial genera ALR-transformed abundances and 1901 microbial gene ALR-transformed abundances as predictor variables, which had variable importance in projection (VIP) score higher or equal to 1. Cells coloured blue and orange represent positive and negative regression coefficients, respectively.

We investigated the importance (VIP) of microbial genera and genes previously identified as biomarkers for the host performance traits, to evaluate whether these were still highly associated with each trait using the longitudinal microbiome data of the present study (Table 6). Microbial genera and genes biomarkers extracted from Roehe et al. (2016), Martínez-Álvaro et al. (2020), and Lima et al. (2019) that met our abundance/zeros criteria (total of 123 biomarkers) were included in these analyses. The results showed that whereas some of these biomarkers had decreased importance in PLS predictions based on their abundances at earlier timepoints, when those predictions were based on end test and slaughter data, they were highly concordant with previous studies, with 81 of 123 biomarkers had VIP ≥ 0.8 in each of these 6 timepoints. The timepoint showing the lowest agreement with previous works was at mid test, in which the PLS models resulted in 66 and 26 biomarkers with VIP ≥ 0.8 and VIP < 0.5, respectively.

Table 6. *Importance (VIP) of microbial genera and genes previously identified as biomarkers for host performance and methane emissions traits in partial least squares analyses based on our microbiome data generated from 6 timepoints.*

Trait	Study	Microbial genus/ gene	Pre-additive VIP	Start test VIP	Mid test VIP	End test VIP	Chamber VIP	Slaughter VIP
CH_4 yield	MA	*Pochonia*	0.84	1.13	1.38	1.25	1.55	1.59
		Tremella	1.27	0.78	1.26	1.08	1.15	1.39
		Niastella	0.94	0.79	0.83	1.34	0.83	1.24
		Bacillus	0.84	0.94	1.09	1.02	1.28	0.74
		Fomitiporia	0.57	0.94	1.54	0.99	1.55	1.08
		Selenomonas	0.51	1.29	1.67	1.93	1.42	1.27
		Tolumonas	0.94	1.78	0.89	0.68	1.30	1.34
		Leclercia	0.98	1.83	1.08	0.46	0.83	1.40
		Moraxella	1.17	1.81	0.90	0.46	0.96	1.41
		Fibrobacter	0.52	0.74	1.31	1.16	1.17	1.46
		Prevotella	1.16	1.06	1.17	0.91	0.76	0.63
		Butyrivibrio	0.99	0.45	0.79	0.62	1.05	1.60
		Salinibacter	1.12	0.73	0.48	0.91	0.77	1.19
		Alloactinosynnema	0.84	0.74	0.73	0.89	0.29	0.69
		Sediminispirochaeta	0.76	0.38	0.70	0.91	0.84	0.80
FCR	L	GLO1 (K01759)	0.85	1.16	0.42	1.17	1.26	0.93
		lctP (K03303)	1.18	0.65	1.08	1.65	1.57	0.87
		nusA (K02600)	1.00	1.56	1.49	1.06	0.64	1.07
		xylE (K08138)	0.34	0.92	1.52	0.87	0.22	0.91
		idnO (K00046)	0.80	0.55	1.19	1.68	1.20	1.31
		galK (K00849)	0.35	0.96	1.31	1.43	0.74	0.81
		gcvH (K02437)	1.50	1.01	0.69	0.65	1.02	1.23
		infA (K02518)	1.04	0.51	1.46	0.97	1.36	0.58
		punA (K03783)	1.16	0.86	0.66	0.70	0.69	1.01
		hupB (K03530)	0.92	0.36	1.77	0.42	1.23	0.53
		lpxA (K00677)	0.89	0.46	0.38	0.58	0.06	1.05
		murD (K01925)	1.78	1.12	0.49	0.40	0.73	0.75
		uidA (K01195)	0.31	0.77	0.16	0.79	0.43	1.14
		aguA (K01235)	0.50	0.47	0.27	0.20	0.99	0.56
	R	*murC* (K01924)	1.18	1.33	1.24	0.93	0.92	0.90
		recD (K03581)	1.12	1.08	0.96	1.10	0.85	1.23
		uvrD (K03657)	1.11	1.47	1.19	1.14	0.99	0.94
		dnaA (K02313)	1.02	1.03	0.97	1.01	0.95	1.03
		DARS2 (K01876)	1.12	1.50	1.01	1.20	0.93	0.83
		gidB (K03501)	0.99	1.03	2.06	0.88	0.14	1.26
		dnaX (K02343)	1.02	1.33	0.76	0.98	0.81	0.89
		RP-L30 (K02907)	1.23	0.51	1.28	1.51	1.56	0.88
		tyrA2 (K04517)	1.09	0.98	1.23	1.02	0.49	1.21
		hisA (K01814)	0.89	0.91	0.81	1.52	1.43	0.45
		pepF (K08602)	1.09	1.17	0.97	1.06	0.77	1.27
		MMAA (K07588)	1.08	0.89	0.69	1.22	1.05	1.19
		thiD (K00941)	1.36	1.06	1.26	1.39	0.27	1.20

		dgt (K01129)	0.99	0.82	0.87	1.35	0.78	0.76
		recN (K03631)	0.87	1.54	0.74	0.91	0.74	0.88
		K07139	1.08	1.83	0.78	1.04	0.63	1.11
		clpP (K01358)	1.18	0.92	0.60	0.50	1.10	1.16
		iorA (K00179)	0.86	0.97	0.51	1.10	1.82	0.57
		pdxK (K00868)	0.53	1.23	0.92	1.56	1.03	0.50
		TSTA3 (K02377)	0.87	0.60	1.17	1.14	0.29	1.48
		psd (K01613)	1.14	0.50	0.66	1.00	0.39	1.00
		ssb (K03111)	1.43	1.46	0.48	0.34	0.87	0.66
		ribD (K11752)	0.90	1.06	0.21	0.97	0.59	0.32
		ftsX (K09811)	2.31	1.08	0.25	0.46	0.72	0.92
		murE (K01928)	0.83	1.70	0.69	1.20	0.04	0.68
		galE (K01784)	0.86	1.98	0.48	0.80	0.66	1.05
		rnfC (K03615)	0.33	1.10	1.33	1.09	0.80	0.75
		murB (K00075)	0.87	1.55	0.72	0.20	0.68	0.88
		nadB (K00278)	0.99	0.79	0.19	0.22	0.96	0.88
		K06921	1.12	0.84	0.19	0.69	0.82	0.55
		trpD (K00766)	0.84	1.46	0.58	0.80	1.24	0.57
		E3.1.3.48 (K01104)	0.71	0.75	1.31	0.53	0.73	0.91
		comEB (K01493)	0.69	0.76	0.85	0.44	1.34	0.65
		cca (K00974)	0.52	0.74	0.50	0.99	0.21	1.12
		nudC (K03426)	0.90	1.18	0.70	0.54	0.46	0.02
		tgt (K00773)	0.58	0.72	0.39	0.73	0.43	0.99
		cbiQ (K02008)	0.70	0.75	0.47	0.67	1.34	0.47
		K07082	0.26	0.57	1.00	0.57	0.65	0.40
		fucI (K01818)	0.31	1.05	0.68	0.68	0.31	0.08
		uidA (K01195)	0.31	0.77	0.16	0.79	0.43	1.14
ADG	L	*gcvH* (K02437)	1.60	1.10	1.31	0.95	0.95	1.20
		paaK (K01912)	0.57	1.45	0.92	0.91	1.43	0.89
		RP-L36 (K02919)	0.93	0.90	1.92	1.23	1.43	0.53
		ATPF1D (K02113)	0.42	1.10	0.86	0.96	1.39	1.79
		gcvPB (K00283)	0.93	1.59	0.94	0.85	0.99	0.80
		nusA (K02600)	0.84	1.45	1.17	1.05	0.44	1.23
		infA (K02518)	1.01	0.91	1.00	0.90	1.41	0.55
		amiABC (K01448)	1.00	1.03	0.48	1.03	0.46	1.00
		asd (K00133)	0.69	1.58	1.80	0.95	1.30	0.68
		slyD (K03775)	1.26	0.73	1.63	1.05	1.67	0.69
		hupB (K03530)	1.45	1.05	1.49	0.41	1.44	0.39
		RP-L17 (K02879)	1.04	1.54	0.47	0.74	0.88	0.77
		murD (K01925)	2.19	1.02	0.72	0.29	0.72	0.63
DFI	L	*tolC* (K12340)	1.54	1.09	0.98	1.40	1.29	0.83
		rpoB (K03043)	0.56	0.99	1.21	1.37	0.90	1.02
		ehbD (K14113)	0.23	1.46	1.38	1.00	1.11	0.92
		glnB (K04751)	0.72	1.35	0.79	1.03	1.33	1.36
		yajC (K03210)	0.32	1.19	1.17	0.69	0.84	1.30
		rnc (K03685)	0.50	0.94	0.75	1.00	0.96	1.18
		mrdA (K05515)	0.21	0.76	0.77	0.88	1.54	0.95

		rfbG (K01709)	1.48	0.93	0.38	0.67	0.93	0.76
		rdgB (K02428)	0.72	0.39	0.46	1.05	0.86	1.18
		INO1 (K01858)	1.14	0.05	1.26	1.23	0.29	0.76
		bglX (K05349)	1.06	0.45	0.75	1.16	0.65	1.05
		rluB (K06178)	0.97	0.52	0.41	0.55	0.75	0.71
		nusB (K03625)	0.45	0.43	0.46	0.54	0.40	1.58
		rfbF (K00978)	1.05	0.64	0.36	0.44	0.47	0.60
RFI	L	*zntA* (K01534)	1.02	1.41	1.15	1.03	1.26	1.20
		dacC (K07258)	1.37	1.55	0.82	1.19	1.87	0.98
		yrbG (K07301)	0.85	1.09	1.15	1.43	0.93	1.06
		ehbD (K14113)	0.88	1.29	1.28	0.96	1.00	0.99
		cobL-cbiET (K00595)	1.43	0.44	1.09	0.93	0.97	1.49
		cbiN (K02009)	0.75	1.32	1.38	0.87	1.02	1.00
		cobD (K04720)	0.69	0.64	1.02	1.49	0.90	1.02
		rfbG (K01709)	0.87	0.78	0.32	1.05	0.88	1.51
		rfbF (K00978)	0.54	0.43	0.26	0.91	0.94	1.57
CH_4 yield	MA	LDH (K00016)	1.02	1.20	1.26	1.09	0.80	1.38
		BCP (K03564)	0.89	1.09	1.24	1.26	1.21	1.07
		metA (K00651)	0.36	1.22	0.84	1.14	1.12	1.40
		livM (K01998)	0.63	1.13	1.44	1.16	1.15	1.25
		K09702	0.74	0.89	1.32	0.43	0.83	1.45
		E1.1.1.219 (K00091)	0.61	1.16	0.79	0.85	1.34	0.48
		xynA (K01181)	0.44	1.07	0.90	1.41	0.36	0.26
		nifB (K02585)	0.22	0.80	0.50	0.77	0.99	0.86
		kbl (K00639)	0.46	0.93	0.72	0.42	1.40	0.18
		nusA (K02600)	0.18	0.65	0.74	1.19	0.78	0.92
		ramA (K05989)	0.75	0.66	0.67	0.74	1.10	1.37
		K06950	0.83	0.65	0.77	0.90	0.72	0.56
		PK (K00873)	0.78	0.78	0.80	0.74	0.56	1.37
		ABC.SN.A (K02049)	0.81	0.68	0.24	0.63	0.52	0.52
	R	*frhB* (K00441)	0.64	1.22	0.94	0.75	0.86	1.51
		fdhB (K00125)	1.47	0.64	0.82	1.20	0.68	0.42
		porB (K00170)	0.52	0.52	0.56	1.00	0.90	0.60
		porA (K00169)	0.27	0.44	0.51	0.55	0.69	0.16

VIP refers to variable importance in projection, obtained from partial least squares prediction models for prediction of methane yield (CH_4 yield; g/kg dry matter intake), feed conversion ratio (FCR), average daily gain (ADG), daily feed intake (DFI), and residual feed intake (RFI). MA, L, and R refer to the manuscripts identified as Martinez-Alvaro et al. (2020), Lima et al. (2019), and Roehe et al. (2016), respectively. Cells coloured in green, and orange represent VIP ≥ 0.8 and VIP < 0.5, respectively.

4.4.4 Evaluation of the stability of the association between host-genomically influenced microbial genes and the estimated breeding values of host performance traits

Microbial genes found to be moderately to highly heritable (with mean heritability between 0.16 and 0.54) in the study of Martínez-Álvaro, et al. (2021) were used as explanatory variables in PLS models, separately analysed for each timepoint, to predict EBVs of host performance traits FCR, ADG, DFI, RFI, and CH_4 yield.

Table 7. *Proportion of explained variance of estimated breeding values of performance and methane emissions traits predicted by ALR-transformed abundances of moderately to highly heritable microbial genes separately analysed for each of the 6 timepoint*

Trait	FCR		ADG		DFI		RFI		CH_4 yield	
Timepoint	X (%)	Y (%)	X (%)	Y (%)	X (%)	Y (%)	X (%)	Y (%)	X (%)	Y (%)
Pre-additive	16.82	81.57	33.83	77.31	47.49	78.86	29.09	69.71	25.38	70.33
Start test	35.38	65.24	35.21	64.10	25.38	75.01	31.34	57.45	33.19	70.44
Mid test	35.77	63.98	23.37	66.07	39.32	71.03	38.58	59.28	36.08	83.11
End test	32.77	61.93	38.18	72.79	41.56	58.42	36.37	60.22	27.57	75.44
Chamber	39.98	67.11	40.53	83.27	27.23	79.20	37.90	78.43	27.66	60.59
Slaughter	23.13	70.87	24.30	71.94	24.72	76.64	22.20	78.88	26.17	64.41

FCR, ADG, DFI, RFI, CH_4 yield, and daily CH_4 emissions refer to feed conversion ratio, average daily weight gain, daily feed intake, residual feed intake, methane yield (g/Kg DMI) and daily methane emissions (g/day) respectively. X (%) and Y (%) refer to percentage variation in the metagenomic dataset and the estimated breeding value of the trait, respectively, obtained from Partial Least Squares prediction models using 2 components and including a total of 443 microbial gene ALR-transformed abundances. The shaded cells represent the timepoint that explained the highest variance of each trait.

The PLS analyses using separately the ALR-As of 443 moderately to highly heritable microbial genes at each timepoint to predict EBVs of each trait revealed that a high proportion of the EBVs of each trait was explained by the microbial genes throughout all timepoints, with averages of 68.5 ± 7.1%, 72.6 ± 7.1%, 73.2 ± 7.8%, 67.3 ± 9.8%, and 70.7 ± 8.0%, respectively for FCR, ADG, DFI, RFI, and CH_4 yield, respectively (Table 7).

A total of 28, 37, 34, 56, and 42 microbial genes were found to be significantly (based on their Jack-Knife interval not containing 0) associated to the EBVs of FCR, ADG, DFI, RFI, and CH_4, respectively, in all timepoints. From these, 7, 17, 12, 28, and 17 were negatively, whereas 8, 8, 5, 15, and 6 were positively

associated with FCR, ADG, DFI, RFI, and CH_4, respectively, in at least 5 timepoints.

Table 8. *Number of highly to moderately heritable microbial genes important to explain the variance of the estimated breeding values (EBV) of each performance and methane emissions trait, and percentage of these that were also important in slaughter samples*

Trait	FCR EBV		ADG EBV		DFI EBV		RFI EBV		CH_4 yield EBV	
Timepoint	Sig (JK)	SL (%)	Sig (JK)	SL (%)	Sig (JK)	SL (%)	Sig (JK)	SL (%)	Sig (JK)	SL (%)
Pre-additive	231	52.2	261	54.6	275	58.5	299	61.5	259	59.6
Start test	292	66.4	286	59.7	258	55.1	265	51.4	303	67.3
Mid test	262	58.1	275	57.1	297	57.7	320	65.5	308	69.2
End test	314	67.6	290	63.4	286	62.1	248	50.7	293	61.2
Chamber	253	57.3	253	55.3	241	51.8	280	58.3	250	51.5
Slaughter	253	100.0	273	100.0	272	100.0	278	100.0	260	100.0

FCR, ADG, DFI, RFI, and CH_4 yield refer to feed conversion ratio, average daily weight gain, daily feed intake, residual feed intake, and methane yield (g/Kg DMI), respectively. Sig (JK) represents the number of microbial genes significantly associated with the trait (based on Jack-Knife interval) in partial least squares prediction models (PLS) based on data generated from each timepoint. SL (%) was calculated as the percentage of microbial genes also significant for prediction of those traits at slaughter. PLS prediction models were calculated using 2 components and including a total of 443 microbial gene ALR-transformed abundances. Shaded cells represent the highest SL (%) for each trait.

Microbial genes with moderate to high heritability strongly associated with the performance traits based on slaughter-generated data were also found to be strongly associated with the performance traits at earlier timepoints (Table 8). For example, 67.6%, 63.4%, 62.1%, 50.7%, and 61.2% of the end test-generated microbial gene ALR-As significantly associated with FCR, ADG, DFI, RFI, and CH_4 yield, were also identified as significant when PLS models were based on slaughter-generated microbiome data.

Table 9. *Number of highly to moderately heritable microbial genes significantly associated to the estimated breeding values of host performance and methane emissions traits throughout the finishing phase, and percentage of these that were also important in slaughter-generated samples*

Association with the trait	Trait	FCR EBV		ADG EBV		DFI EBV		RFI EBV		CH_4 yield EBV	
	Timepoint	Sig (JK)	SL (%)	Sig (JK)	SL (%)	Sig (JK)	SL (%)	Sig (JK)	SL (%)	Sig (JK)	SL (%)
[1]Significant positive	Pre-additive	174	16.7	102	47.7	150	47.5	65	57.7	186	32.9
	Start test	162	23.7	104	57.8	104	48.5	190	22.7	84	55.7
	Mid test	145	29.0	102	37.6	94	43.4	110	34.0	78	54.4
	End test	154	18.4	82	39.5	68	58.6	63	55.7	47	65.8
	Chamber	64	75.4	180	35.8	86	57.6	78	44.3	83	43.0
	Slaughter	114	100.0	109	100.0	99	100.0	97	100.0	79	100.0
[1]Significant negative	Pre-additive	79	43.2	171	72.6	122	43.4	213	86.7	74	26.5
	Start test	91	46.0	169	64.6	168	69.4	88	36.5	176	72.9
	Mid test	108	54.0	171	79.3	178	78.3	168	74.6	182	76.8
	End test	99	56.1	191	90.2	204	84.4	215	89.0	213	89.0
	Chamber	189	74.1	93	32.9	186	74.6	200	86.7	177	79.0
	Slaughter	139	100.0	164	100.0	173	100.0	181	100.0	181	100.0

FCR, ADG, DFI, RFI, and CH_4 yield refer to feed conversion ratio, average daily weight gain, daily feed intake, residual feed intake, and methane yield (g/Kg DMI), respectively. [1]Significance was assessed by Jack-Knife procedure of regression coefficients obtained through leave-one-out cross-validation and the mean ± 2 standard deviations interval not containing 0. Sig (JK) represents the number of microbial genes significantly associated with the trait (based on Jack-Knife procedure) in partial least squares models (PLS) based on slaughter-generated data, that had a positive or negative association with the trait in each timepoint. SL (%) was calculated as the percentage of microbial genes significantly (positively or negatively) associated with the host traits in previous timepoints that were also positively or negatively associated with the traits for prediction of those traits at slaughter. Shaded cells represent the highest SL (%) for each trait.

In addition, several of these associations were found to maintain the signal of their relationship (i.e., the signal of the regression coefficient of the first component in the PLS, Table 9). For example, 253 microbial gene ALR-As based on rumen samples at slaughter were significantly associated with FCR in the PLS model. Using microbiome data obtained shortly after leaving the respiration chamber, 64 and 189 microbial gene ALR-As had positive and negative association, respectively, with FCR, constituting 75.4% and 74.1% of the 114 and 139 microbial gene ALR-As at slaughter that had positive and negative associations with the trait, respectively. The agreement between the associations was particularly high regarding the microbial genes with negative associations with the traits based on PLS models microbiome data generated from slaughter samples. For example, 56.1%, 90.2%, 84.4%, 89.0%, and

89.0% of the microbial genes showing negative associations with FCR, ADG, DFI, RFI, and CH_4 yield, respectively, in PLS models based on microbiome data at end test, also showed negative association with the traits when models were based on microbiome data at slaughter.

4.5 Discussion

Based on all results, the rumen microbiome at microbial genera and genes level showed high temporal stability over the entire finishing phase of beef steers based on longitudinal ruminal samples taken at 6 timepoints, approximately within one month of each other.

At the microbial genes level, diversity was highly stable throughout the finishing phase except for samples taken at slaughter, which showed significantly higher evenness than those taken at mid test. In agreement, Qiu et al. (2019) investigated the taxonomic alpha diversity of rumen samples throughout 3 months of adaptation to new diets in finishing steers and observed no significant differences between timepoints (no post-slaughter samples were investigated). Additionally, PERMANOVA and NMDS plot of the samples based on Bray-Curtis dissimilarities indicates that even when the differences were significant, the effect of timepoint was not strong, because it accounts only for a maximum of 14% of the variation between microbiome compositions. Considering that the rumen liquid digesta has been suggested to harbour different communities from those found in the rumen epithelium (Sadet et al., 2007; Mao et al., 2015), the differences between preceding timepoints and slaughter may be due to the sample collection methods applied; whereas before slaughter, the samples were obtained through a nasal tube (liquid digesta samples), after slaughter the samples were collected directly from the open rumen at the abattoir, which results in well mixed samples (liquid digesta and epithelium samples). In a previous study by Snelling et al. (2019), Bray-Curtis dissimilarities were shown to significantly differ between Operational Taxonomic Units (OTUs) datasets generated from samples taken from concentrate-fed animals at pre-additive and 25 days after the pre-additive timepoint (adaptation timepoint, not included in the present

study), but no differences were observed regarding forage-fed animals. Considering that these animals had been adapted to concentrate diet for a period or 4 weeks, and prior to that they were forage-fed, this significant difference could be associated to the adaptation of the microbiome to the new diet, rather than to time. In agreement, in a study on finishing beef cattle, Clemmons et al. (2019) showed that the rumen microbiome started to shift to a stable microbial community 4 weeks after the change in diet, and continued to stabilize over the next 5 weeks, and recommended an adaptation period of at least 8 weeks following the transition from a forage-based to a concentrate-based diet.

The ALR-As of the microbial genera and genes observed at each timepoint were compared using linear mixed models, revealing that the abundance of 99% of the microbial genera and of the microbial genes were maintained throughout the 6 timepoints, suggesting that the microbiome has a high level of stability during the finishing phase of steers.

Since the performance traits are calculated based on measurements taken during the RFI testing period, it would be expectable that the microbiota and microbial genes datasets generated from samples obtained during the performance testing period (i.e., start, mid, and end test) explained the highest variance in each of these performance traits, and some of our results agreed with these expectations, particularly when the analyses were based on the abundance of microbial genera. For example, the highest proportion of variation in FCR, ADG, and DFI (70.19%, 67.77%, and 78.33%, respectively) was explained by microbial genera generated from start test samples. Overall, the explained variances for each trait were substantially high throughout all timepoints, particularly when using microbial genes as predictor variables, and showed low variation throughout the timepoints, suggesting that the rumen microbiome collected at any timepoint during the finishing phase is highly associated to performance and methane emissions traits including appetite, growth, feed conversion efficiency, daily methane emissions, and methane yield.

The correlations between microbiome datasets generated from samples taken at different timepoints showed that at both microbial genera and genes levels, the highest correlation to the compositions generated from slaughter samples were observed for samples taken after the animals left the respiration chambers; additionally, all pairwise correlations were high overall, with the lowest being 70% and 81% in the cases of microbial genera and genes, respectively.

The influence of the nitrate and oil-based additives on the daily methane emissions and yield of these same animals was investigated by Troy et al. (2015), who reported reductions of 17% and 7.5% in methane yield from animals fed forage diet with nitrate and oil additives, respectively, but no effect was reported for additives when animals were fed concentrate.

We also analysed the temporal stability of moderately to highly heritable microbial genes, in association with the EBVs of FCR, ADG, DFI, RFI, and CH_4 yield. The results showed that the ALR-As of microbial genes generated from any timepoint explained 67% or higher proportion of the variation of the EBVs. Additionally, 51% or more of the microbial genes generated from slaughter were also significantly associated with each trait when models were based on data generated from previous timepoints, suggesting that heritable microbial genes and their associations with the traits have high temporal stability.

The rumen microbiome has been shown to significantly shift over the early life of calves. O'Hara et al. (2020) investigated the rumen microbial dynamics of beef calves based on 16S rRNA gene sequencing (i.e., bacterial and archaea communities) and showed that age had a significant impact on the rumen microbiota composition, which became stable when calves were 21 days old, and suggested that this 3-week window could be the most advantageous for microbiota manipulation. Additionally, Jami et al. (2013) showed the microbiome of young calves undergoes drastic changes, particularly associated to the rumen maturation, with progressive depletion of aerobic and facultative aerobic groups, establishment of cellulolytic microbes long before

exposure to plant material, and increased taxonomic diversity with age. In adult animals, however, the expectation is that, in the absence of major perturbations (i.e., diet changes), the microbiome will remain relatively stable, not only in bovines but also other animals. Previous authors have shown this tendency; for example, Schloissnig et al. (2013) investigated human gut microbial genomic variation from 207 humans, and showed that individual-specific variation patters were stable over time, Snelling et al. (2019) investigated the rumen microbiota (resolved based on the 16S rRNA gene) in steers, and concluded that a single sample may be reasonably representative of the microbial communities in the rumen, in agreement with the analyses presented in this study.

Most often, the association of microbial genera and genes with host performance and methane emissions traits is based on microbiome profiles generated from slaughter samples (Guan et al., 2008; Mao et al., 2015; Roehe et al., 2016; Li and Guan, 2017; Lima et al., 2019; Martínez-Álvaro et al., 2021a), but whether the slaughter samples are suitable representatives of the microbiome composition in previous moments of hosts' life is still unclear. We have analysed biomarkers previously identified as strong predictors of performance and methane emissions traits in studies based on slaughter samples; our results suggest that the rumen microbiome has such stability that the previously identified important biomarkers would still be considered important if these analyses had been based on data generated from any timepoint other than slaughter. Our results confirm the suitability of these slaughter-generated biomarkers, particularly when considering that these previous studies applied different statistical procedures (e.g., correlation networks), on data transformed by different methods (e.g., relative abundances), and that the identification of microbial genera and genes was performed based on different reference databases (e.g., through update of KEGG databases).

4.6 Conclusions

Overall, our study showed that the rumen microbiome is remarkably stable throughout the finishing phase of beef cattle. Alpha and beta diversity compared between timepoints were overall stable at both microbial genera and genes levels. Additionally, most ALR-As of microbial genera and genes within animal were maintained throughout this period. The predictability (measured as explained variation of the traits) of host performance and methane emissions traits using microbiome information collected at each of the 6 different timepoints was strong throughout the entire finishing phase. The high stability of microbial ALR-As of microbial genera and genes could be one of the reasons for the high predictability of host traits based on ALR-As throughout all 6 timepoints. The microbiome datasets that led to the strongest associations with the host performance and methane emissions traits were the ALR-As of microbial genera generated from samples collected at start test, and those of microbial genes generated from samples collected at pre-additive. Overall, the use of microbial genes as predictors leads to higher explained variance of host performance and methane emissions traits. Daily CH_4 emissions and CH_4 yield were, from the analysed traits, the ones most strongly predicted based on microbial genera, and microbial genes, respectively. Our analyses of previously identified biomarkers of host performance and methane emissions traits based on slaughter-generated data underlines the stability between microbiome information throughout the finishing phase. The ALR-As of host-genomically influenced microbial genes generated from samples taken after the animals left the respiration chambers in which they were measured for methane emissions were the most strongly associated with EBVs of host performance and methane emissions traits. Overall, host-genomically influenced microbial genes were strongly associated with the EBVs of all host performance and methane emissions traits, but the association was particularly stronger between host-genomically influenced microbial genes and the EBV of DFI. Our results indicate that microbiome profiles generated from samples collected at any timepoint can be used to accurately predict host performance and methane emissions traits, suggesting

that microbiome profiles generated from timepoints as earlier as that at pre-additive can be applied to predict breeding values of the analysed traits.

4.7 References

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi:10.1093/bioinformatics/btz859.

Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C. A., Watson, M., and Roehe, R. (2020). Identification of microbial genetic capacities and potential mechanisms within the rumen microbiome explaining differences in beef cattle feed efficiency. *Front. Microbiol.* 11. doi:10.3389/fmicb.2020.01229.

Auffret, M. D., Stewart, R., Dewhurst, R. J., Duthie, C.-A., Rooke, J. A., Wallace, R. J., et al. (2018). Identification, comparison, and validation of robust rumen microbial biomarkers for methane emissions using diverse Bos taurus breeds and basal diets. *Front. Microbiol.* 8, 1–15. doi:10.3389/fmicb.2017.02642.

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67. doi:10.18637/jss.v067.i01.

Cao, K.-A. Le, Rohart, F., Gonzalez, I., Dejean, S., Abadi, A., Gautier, B., et al. (2020). Package 'mixOmics.'

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560.

Clemmons, B. A., Martino, C., Schneider, L. G., Lefler, J., Embree, M. M., and Myer, P. R. (2019). Temporal stability of the ruminal bacterial communities in beef steers. *Sci. Rep.* 9, 1–8. doi:10.1038/s41598-019-45995-2.

Danielsson, R., Dicksved, J., Sun, L., Gonda, H., Müller, B., Schnürer, A., et al. (2017). Methane production in dairy cows correlates with rumen methanogenic and bacterial community structure. *Front. Microbiol.* 8, 1–15. doi:10.3389/fmicb.2017.00226.

Duthie, C.-A., Rooke, J. A., Troy, S., Hyslop, J. J., Ross, D. W., Waterhouse, A., et al. (2016). Impact of adding nitrate or increasing the lipid content of two contrasting diets on blood methaemoglobin and performance of two breeds of finishing beef steers. *Animal* 10, 786–795. doi:10.1017/S1751731115002657.

Fox, J., and Weisberg, S. (2019). *An R Companion to Applied Regression*. 3rd ed. , ed. Sage Thousand Oaks CA.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* 8, 1–6. doi:10.3389/fmicb.2017.02224.

Greenacre, M., Martínez-Álvaro, M., and Blasco, A. (2021). Compositional data analysis of microbiome and any-omics datasets : a revalidation of the additive logratio transformation. doi:doi: https://doi.org/10.1101/2021.05.15.444300.

Guan, L. L., Nkrumah, J. D., Basarab, J. A., and Moore, S. S. (2008). Linkage of microbial ecology to phenotype: Correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol. Lett.* 288, 85–91. doi:10.1111/j.1574-6968.2008.01343.x.

Huws, S. A., Edwards, J. E., Creevey, C. J., Stevens, P. R., Lin, W., Girdwood, S. E., et al. (2016). Temporal dynamics of the metabolically active rumen bacteria colonizing fresh perennial ryegrass. *FEMS Microbiol. Ecol.* 92, 1–12. doi:10.1093/femsec/fiv137.

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *Nat. Commun.* 11, 1–8. Available at: http://dx.doi.org/10.1016/B978-0-12-407863-5.00023-X%5Cnhttp://www.nature.com/doifinder/10.1038/ismej.2009.79%5Cnhttp://www.nature.com/doifinder/10.1038/nature09916%5Cnhttp://dx.doi.org/10.1038/srep25982%5Cnhttp://dx.doi.org/10.1038/ismej.2010.144%5Cnhttp.

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

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1016/j.meegid.2016.07.022.

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

Li, F., and Guan, L. L. (2017). Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Appl. Environ. Microbiol.* 83. doi:10.1128/aem.00061-17.

Lima, J., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C.-A., Snelling, T. J., et al. (2019). Identification of rumen microbial genes involved in pathways linked to appetite, growth, and feed conversion efficiency in cattle. *Front. Genet.* 10:701. doi:10.3389/fgene.2019.00701.

Mao, S., Zhang, M., Liu, J., and Zhu, W. (2015). Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: Membership and potential function. *Sci. Rep.* 5, 1–14. doi:10.1038/srep16116.

Martínez-Álvaro, M., Auffret, M. D., Duthie, C.-A., Dewhurst, R. J., Cleveland, M., Watson, M., et al. (2021a). Bovine host genome acts on specific metabolism, communication and genetic processes of rumen microbes host-genomically linked to methane emissions. *Res. Sq. - Prepr.*, 0–37.

Martínez-Álvaro, M., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., et al. (2020). Identification of complex rumen microbiome interaction within diverse functional niches as mechanisms affecting the variation of methane emissions in bovine. *Front. Microbiol.* 11:659. doi:10.3389/fmicb.2020.00659.

Martínez-Álvaro, M., Mattock, J., Auffret, M., Weng, Z., Duthie, C.-A., Dewhurst, R., et al. (2021b). Microbiome-driven breeding to improve the omega-3 and conjugated linoleic fatty acids in beef with simultaneous reduction in methane emission (submitted). *Microbiome*.

Martínez-Álvaro, M., Zubiri-Gaitán, A., Hernández, P., Greenacre, M., Ferrer, A., and Blasco, A. (2021c). Comprehensive comparison of the cecum microbiome functional core in genetically obese and lean hosts under the same environmental conditions. *Commun. Biol.*

Martinez Arbizu, P. (2020). pairwiseAdonis: Pairwise multilevel comparison using adonis.

Myer, P. R., Smith, T. P. L., Wells, J. E., Kuehn, L. A., and Freetly, H. C. (2015). Rumen microbiome from steers differing in feed efficiency. *PLoS One* 10, e0129174. doi:10.1371/journal.pone.0129174.

O'Hara, E., Kenny, D. A., McGovern, E., Byrne, C. J., McCabe, M. S., Guan, L. L., et al. (2020). Investigating temporal microbial dynamics in the rumen of beef calves raised on two farms during early life. *FEMS Microbiol. Ecol.* 96, 1–16. doi:10.1093/femsec/fiz203.

Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2019). Package 'vegan.'

Pérez, P., and De Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi:10.1534/genetics.114.164442.

Piao, H., Lachman, M., Malfatti, S., Sczyrba, A., Knierim, B., Auer, M., et al. (2014). Temporal dynamics of fibrolytic and methanogenic rumen microorganisms during in situ incubation of switchgrass determined by 16s rRNA gene profiling. *Front. Microbiol.* 5, 1–11. doi:10.3389/fmicb.2014.00307.

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35, 61–65. doi:10.1093/nar/gkl842.

Qiu, Q., Gao, C., Gao, Z., Rahman, M. A. U., He, Y., Cao, B., et al. (2019). Temporal dynamics in rumen bacterial community composition of finishing steers during an adaptation period of three months. *Microorganisms* 7, 1–16. doi:10.3390/microorganisms7100410.

R Core Team (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria Available at: https://www.r-project.org/.

Roehe, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., McKain, N., Ross, D. W., et al. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* 12, e1005846. doi:10.1371/journal.pgen.1005846.

Rooke, J. A., Wallace, R. J., Duthie, C.-A., McKain, N., Souza, S. M. de, Hyslop, J. J., et al. (2014). Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype. *Br. J. Nutr.* 112, 398–407. doi:10.1017/s0007114514000932.

Russell, A., Lenth, V., Buerkner, P., Herve, M., Love, J., Singmann, H., et al. (2021). Package ‘ emmeans ’ R topics documented : 34, 216–221. doi:10.1080/00031305.1980.10483031>.License.

Sadet, S., Martin, C., Meunier, B., and Morgavi, D. P. (2007). PCR-DGGE analysis reveals a distinct diversity in the bacterial population attached to the rumen epithelium. *Animal* 1, 939–944. doi:10.1017/S1751731107000304.

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., et al. (2013). Genomic variation landscape of the human gut microbiome. *Nature* 493, 45–50. doi:10.1038/nature11711.

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

Snelling, T. J., Auffret, M. D., Duthie, C., Stewart, R. D., Watson, M., Dewhurst, R. J., et al. (2019). Temporal stability of the rumen microbiota in beef cattle, and response to diet and supplements. *Anim. Microbiome* 1, 1–14. doi:10.1186/s42523-019-0018-y.

Troy, S. M., Duthie, C.-A., Hyslop, J. J., Roehe, R., Ross, D. W., Wallace, R. J., et al. (2015). Effectiveness of nitrate addition and increased oil content as methane mitigation strategies for beef cattle fed two contrasting basal diets. *J. Anim. Sci.* 93, 1815–1823. doi:10.2527/jas.2014-8688.

Wallace, R. J., Rooke, J. A., Duthie, C.-A., Hyslop, J. J., Ross, D. W., McKain, N., et al. (2014). Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. *Sci. Rep.* 4. doi:10.1038/srep05892.

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

Chapter 5 Rumen microbiome profiles of dairy cattle are affected by the presence of, and vaccination against, the abomasal parasitic nematode *Ostertagia ostertagi*

5.1 Abstract

Parasitism of ruminants by nematodes such as *Ostertagia ostertagi* is reported to negatively affect performance and welfare, leading to inappetence, impaired weight gain, and sometimes death, but whether the rumen microbiome is also affected is still unclear. In this research, we investigated the influence of the abomasal nematode *Ostertagia ostertagi* on the rumen microbiome, at the taxonomic (microbial genera) and functional (microbial genes) levels. Two groups of 10 calves (balanced for weight and breed) were subject to an infection challenge consisting of oral administration of 1000 infectious L3-stage larvae for 25 days. Each group had previously received either a native vaccine against *O. ostertagi* or a co-adjuvant-only injection (positive control). From the infected group, a total of 8 animals were identified based on their high (CHE) and low (CLE) cumulative faecal egg counts (cFEC) and 4 animals were chosen from the vaccinated (VAC) group based on their average cFEC. These, together with 4 unvaccinated and uninfected animals (UNF, negative control), had their rumen digesta sampled post-mortem (20 days after the infection challenge finished). Shotgun metagenomic sequencing was used to resolve the samples into microbiota and metagenomic compositions. The rumen microbiomes of CHE, CLE, and VAC were compared in a pairwise manner against those of UNF using partial least squares discriminant analyses. Our analyses identified 294, 314, and 330 microbial genera important for the CHE vs. UNF, CLE vs. UNF, and VAC vs. UNF discriminations, respectively. Eighty-six microbial genera were altered in CHE, CLE, and VAC, in comparison to UNF, and included microbial genera previously associated with gastrointestinal parasites in animals, such as *Bacillus* and *Deferribacter*. Additionally, the parasitism by *O. ostertagi* led to increased abundance of

opportunistic pathogens (e.g., *Streptomyces* and *Tsukamurella*). The microbiome profiles of infected (INF) were compared with those of VAC; the results showed that microbial genes enriched in INF were mostly associated with methane metabolism (e.g., *frhD*) and carbon metabolism (e.g., *rbcL*), whereas those enriched in VAC were mostly ABC transporters (e.g., *cbiN*), or involved in quorum sensing (e.g., ABC.SP.S) and the two-component system (e.g., *pilR*). This study reveals that the rumen microbiome profiles of infected, vaccinated, and uninfected cattle differ substantially due to the infection by abomasal parasite *O. ostertagi* and could be used as biomarkers for breeding programmes, probiotics development, and for measuring side effects of vaccination.

5.2 Introduction

Ostertagia ostertagi is an abomasal parasitic nematode of cattle, that negatively impacts the gastrointestinal function of the host. *Ostertagia*'s life cycle consists of a free-living phase and a parasitic phase. The free-living phase starts when eggs develop in deposited faeces into first-stage larvae (L1), hatch and moult into second-stage larvae (L2), which develop and moult into infective third-stage larvae (L3). These larvae move onto the herbage, where they are ingested by the grazing animals, starting the parasitic phase of the life cycle (Myers and Taylor, 1989). The L3 larvae lose their protective sheath in the rumen, and pass into the abomasum, where they penetrate the gastric glands and develop into forth-stage larvae (L4) and subsequently to young adults (L5). The L5 leave the gastric glands and continue their maturation in the mucosal surface, and females will produce the eggs in the gastrointestinal tract lumen (Saverwyns, 2008). Ostertagiasis has three clinically distinctive conditions. Type I is a disease of young cattle, characterized by damage to the abomasum gastric glands caused by the L5 exiting the cells, with clinical signs depending on level of infection, ranging from reduced appetite and stunted growth to diarrhoea, rapid weight loss and even death; pre-type II disease occurs when larval development is arrested after entering the gastric glands, and it is usually clinically mild or silent; type II

disease is most common in yearling (older) cattle and is associated with the continuation of the development of previously inhibited larvae, having clinical signs similar to type I disease (Myers and Taylor, 1989).

One of the most economically impactful and widely recognized effects of *Ostertagia* parasitism in cattle is the host's inappetence (Fox et al., 1987, 1989a; Myers and Taylor, 1989; Fox, 1993; Hawkins, 1993; Fox et al., 2002). For example, Fox et al. (1989) reported depressed voluntary feed intake associated with nearly 73% reduction in weight gain, and a 74% decrease in rate of passage of digesta in infected animals, and suggested these signs to be associated with hypergastrinaemia (i.e., increased blood gastrin levels). This association was further confirmed by Fox et al. (2002), with the authors reporting abomasal pH to be positively correlated with blood gastrin levels (and no significant difference in the satiety-associated cholecystokinin levels, Farningham et al., (1993)), suggesting the elevated pH (due to damaged gastric glands) as the main cause for the hypergastrinaemia in infected animals; a positive correlation between blood gastrin and serum pepsinogen levels was also previously reported (Fox et al., 1989b). Serum pepsinogen levels are considered a key parameter to monitor nematode exposure in first-season calves (Charlier et al., 2011), and a positive correlation between the relative pepsinogen concentration and *O. ostertagi* numbers has been found (Berghen et al., 1993; Kenyon and Jackson, 2012). Pepsinogen is the inactive precursor of the aspartic protease pepsin; it is produced in the parietal cells, and it is autocatalytically activated by acidic conditions like those in the abomasum (Muñoz et al., 2004). It has been suggested that pepsinogen increase in the blood is due to increased permeability of the abomasal mucosa, with pepsinogen leaking back into the circulatory system (Murray, 1969).

The inappetence of animals associated with the nematode infection carries obvious negative consequences for productivity. Almería et al. (2009) reported a negative correlation between optical density ratios of milk samples (ODR, proxy for infection level) and milk yield, in temperate climates.

The gastrointestinal tract microbiota is closely associated with the hosts' digestive functional efficiency in many different species, including humans (Cani and Knauf, 2016), pigs (Wellmann et al., 2017; Nowland et al., 2019), chickens (Díaz-Sánchez et al., 2019), and cattle (Shabat et al., 2016). In beef cattle, the rumen microbiome has been associated with feed conversion efficiency and methane emissions (Roehe et al., 2016; Martínez-Álvaro et al., 2020), appetite, and growth (Lima et al., 2019). Given the close associations between the rumen microbiome and performance traits of the host, the investigation of the impact of parasitism by *O. ostertagi* on the rumen microbiome profiles is of great interest as biomarkers for identification of animals resilient to infection and in the development of probiotics to mitigate the impact of infection.

Nematode-parasitic infections have been suggested to lead to increased abundance of potentially pathogenic microorganisms in the gastrointestinal microbiome of the host. For example, Li et al. (2016) reported that although absent from the abomasal microbiome of uninfected goats, pathogenic Gram-negative microorganisms of the *Pasteurellaceae* family were significantly increased in those challenged with *Haemonchus contortus* infectious larvae. Additionally, Lass et al. (2013) showed that co-infecting mice with the pathogen *Bordetella bronchiseptica* promoted the survival of the intestinal nematode *Heligmosomoides polygyrus*.

Immunity to *O. ostertagi* develops only slowly, with cattle remaining susceptible even after long grazing periods. In developed production systems, this parasite is controlled through the application of anthelminthics, and vaccines are being developed. The application of a vaccine against *O. ostertagi* in comparison to the use of anthelminthics is advantageous because it prevents development of potential anthelminthic-resistant parasites. Ideally, a vaccine should build up the animal's immunity, preventing or halting the establishment and consequent damage to the abomasum gastric glands caused by the natural life cycle of this parasite, and therefore avert typical ostertagiasis clinical signs (e.g., inappetence), while not unfavourably

impacting the rumen microbiome. Furthermore, since the microbiome deeply influences the host immune responses, and has been shown to affect vaccine immunogenicity and efficacy (de Jong et al., 2020), understanding the interplay between the rumen microbiome and the vaccine is crucial for the development of efficient vaccines.

The main objective of this study was to understand whether parasitism by *O. ostertagi* impacted the rumen microbiome composition, and, if so, to identify the main changes at the taxonomic level (by, for example, focusing on potentially pathogenic microbes) and at the functional level (by focusing on microbial KEGG genes). We also aimed at understanding whether the vaccine against *O. ostertagi* influenced the rumen microbiome profiles. Subsequently, we focused on the association between the taxonomic and functional infection- and vaccine-derived alterations and investigated them in light of the rumen microbiome complex biological networks, essential for the rumen proper function. Additionally, we investigated how the host animal's resilience to *O. ostertagi* (as assessed by cFEC) influenced taxonomic and functional changes in the rumen microbiome profiles.

5.3 Materials and methods

5.3.1 Ethics statement

Immunizations and parasite challenges of cattle were performed at the Moredun Research Institute (MRI) under Home Office licence 70/7914. Ethical approval was obtained from the MRI Animal Welfare and Ethical Review Body (E12/18). Animals were euthanized at the MRI post-mortem facility.

5.3.2 Animals and experimental procedure

An experimental trial was carried out in 2018 to determine the effect of a native vaccine against *O. ostertagi* in dairy cattle (British Friesian and Norwegian Red).

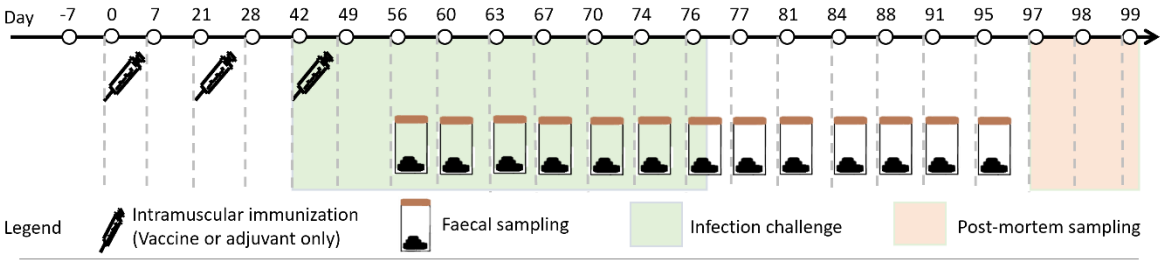


Figure 1. *Schematic representation of the experimental timeline.*

This experiment included 20 male calves of 4-5 months of age. Animals were allocated into 2 groups, balanced for breed and weight. Animals in each group were injected either with the native vaccine or with adjuvant-only (Quil A® (Brenntag Biosector) - 750ug per dose, positive control). Administration of vaccine or adjuvant-only occurred on days 0, 21, and 42 of the experimental trial (Figure 1). The infection challenge started on day 42; animals were orally administered 1000 infectious L3 larvae per day, for 25 days. During the experimental trial, faeces samples were collected at 13 timepoints, and animals were evaluated for their cumulative faecal egg count (cFEC). The cFEC is an indicator of worm fitness, and is used to determine vaccine efficacy against *O. ostertagi* (Meyvis et al., 2007). Based on cFEC distribution per group, a total of 12 animals were identified for post-mortem rumen digesta sampling: 8 animals from the non-vaccinated group were selected based on their high and low cFEC (CHE and CLE, respectively) and 4 animals from the vaccinated group (VAC) were selected from the cFEC-boxplot's second and third quantiles. Additionally, 4 uninfected and unvaccinated calves were kept throughout the experimental trial as negative control (UNF) and were also identified for post-mortem rumen digesta sampling.

5.3.3 Sampling of rumen digesta and whole metagenomic sequencing

Animals included in the experimental trial were slaughtered at the Moredun Research Institute. Sixteen samples of rumen digesta each were collected within 10 minutes after slaughter. Sampling rumen digesta at slaughter results

in well mixed samples. DNA was extracted from the samples following the methodology described by Yu and Morrison (2004).

5.3.4 Identification of abundances of microbial organisms and microbial genes in rumen samples

Illumina TruSeq DNA Nano libraries were prepared from genomic DNA and sequenced on Illumina NovaSeq 6000 systems by Edinburgh Genomics (Edinburgh, UK). Paired-end reads (2 × 150 bp) were generated, resulting in between 8 and 50 GB per sample (between 28 and 165 million paired reads). For taxonomic classification, the sequence reads of the samples were aligned to a database including genomes from the Hungate 1000 Collection (Seshadri et al., 2018) and metagenome-assembled genomes (MAGs) from beef rumen samples (Stewart et al., 2018) using Kraken (Wood and Salzberg, 2014). In total, 1200 genera found in all animals were identified and described as the genus having the highest similarity with the identified microbial genome or MAG. For functional annotation, DIAMOND was used to blast the reads for each sample against the KEGG database (downloaded 15/09/18) (Buchfink et al., 2015). Gene abundance was calculated as the sum of reads mapping to each KEGG orthologue.

5.3.5 Statistical analyses

Alpha diversity was assessed by the number of taxa/microbial genes observed per sample and the adjusted Shannon index (Sobs and H'adj, respectively). Beta diversity was assessed by calculating Bray-Curtis dissimilarity over samples of the same type (BC). Diversity measures were estimated using the vegan package in R studio (Version 1.3.959).

A total of 1200 genera and 8393 microbial genes (i.e., KEGG orthologues, Kanehisa and Goto (2000)) were identified over the 16 samples, and their relative abundances were calculated. Genera and microbial genes that were absent from at least one sample and/or had average relative abundance lower than 0.001% were removed from the datasets, leaving a total of 899 genera and 3124 microbial genes for further analyses. Prior to the statistical analyses,

these datasets were centred-log ratio (CLR) transformed, due to the compositional nature inherent to microbiome data (Greenacre, 2018). Partial Least Squared Discriminant Analysis (PLS-DA) models were applied to evaluate the suitability of the microbiome profiles at microbial genera and microbial genes level (using their CLR-transformed abundances) to discriminate CLE, CHE, and VAC from UNF. Variable importance in projection scores (VIP) were obtained from each PLS-DA, and the threshold of VIP ≥ 1 was used to identify the variables (microbial genera or microbial genes) that most contributed to the discrimination between treatments. We also compared the microbiome profiles of VAC with those derived from infected animals (INF, i.e., CHE and CLE), to better understand the impact of the vaccine. The PLS-DA analyses was performed fitting 2 latent components. These analyses were performed using the 'mixOmics' package in R studio (Version 1.3.959).

A co-abundance network was created using the Graphia Pro software (Dimonaco et al., 2021), based on the CLR-transformed abundances of 3124 microbial genes derived from the UNF animals, using a minimum correlation threshold of r=0.98. Clustering was performed using the Markov clustering method (MCL) available in Graphia Pro, using the granularity value of 2. Information pertaining to the biochemical pathways each microbial gene takes part in (according to KEGG database), and to whether the microbial gene/genus was considered important for the discrimination of CLE, CHE, and VAC from UNF (i.e., VIP ≥ 1 in the PLS-DA models) was included in the network and used in enrichment analyses of the clusters.

5.4 Results

5.4.1 Pairwise comparisons of infected against uninfected animals

Alpha- and beta-diversity indices of CLE, CHE, and VAC were compared with those of UNF in a pairwise manner and revealed no significant differences at either genus or microbial gene levels.

In contrast, PLS-DA comparison of microbiome profiles of CLE with those of UNF revealed significant discrimination at microbial genus and gene levels (Figure 2). From a total of 899 genera, 314 were important (VIP ≥ 1) to discriminate between CLE and UNF microbiomes, of which 126 had lower relative abundance in CLE, including *Candidatus Xiphinematobacter*, *Sphaerochaeta,* and *Enterococcus* (41%, 19%, and 19% decreases, respectively), whereas *Isoptericola*, *Tsukamurella,* and *Aminobacter* were increased in CLE (70%, 64%, and 56% increases, respectively). Regarding the microbial gene-based analysis, from 3124 microbial genes, 1408 showed VIP ≥ 1, e.g., CLE animals had 27%, 17%, and 13% lower relative abundance of *SpsF* (spore coat polysaccharide biosynthesis protein), *modE* (molybdate transport system regulatory protein), and *acpD* (FMN-dependent NADH-azoreductase), respectively, and 95%, and 18% higher relative abundances of *lin* (lincosamide nucleotidyltransferase A/C/D/E), and *thrB* (homoserine kinase), respectively, than UNF.

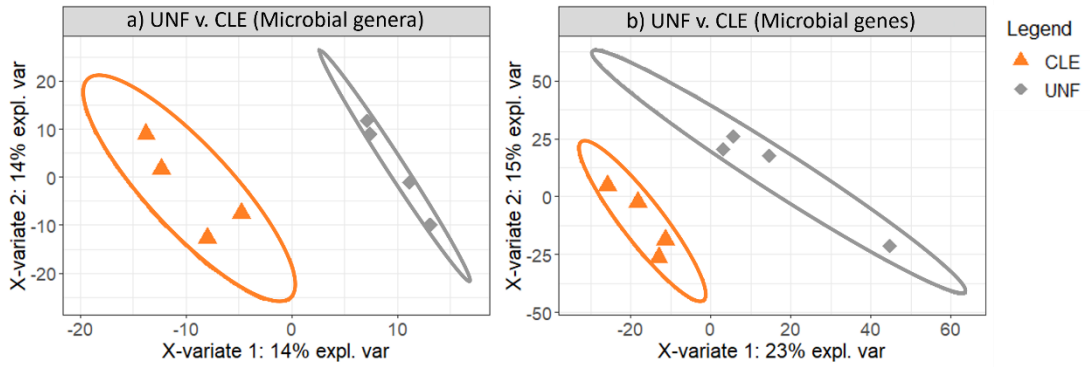


Figure 2. *Partial least squares discriminant analyses (PLS-DA) plots of individuals discriminating between uninfected and infected animals with low cumulative faecal egg count (UNF and CLE, respectively) using PLS-DA based on a) 899 microbial genera and b) 3124 microbial genes (i.e., KEGG level).*

Comparing CHE with UNF in the PLS-DA analysis revealed significant discrimination between treatment groups (Figure 3), with 294 genera and 1060 microbial genes considered important for the discrimination (VIP ≥ 1) in the respective models; for example, *Syntrophomonas*, *Nitratifractor* and *Macrococcus* showed, respectively, 6%, 21%, and 20% lower relative

abundance, whereas *Auricularia*, *Fimbriimonas* and *Candidatus Methanoplasma* had 44%, 45%, and 58% higher relative abundances in CHE in comparison to UNF. For the corresponding comparison of treatment groups at the microbial genes level, *xthA* (exodeoxyribonuclease III), *nifB* (nitrogen fixation protein), and *modE* had respectively 2%, 13%, and 11% decreased abundances, whereas *nfo* (deoxyribonuclease IV), *comFA* (a competence protein), and *spnN* (dTDP-3,4-didehydro-2,6-dideoxy-alpha-D-glucose 3-reductase) showed, respectively, relative abundance increases of 9%, 34%, and 17% in CHE, in comparison to UNF.

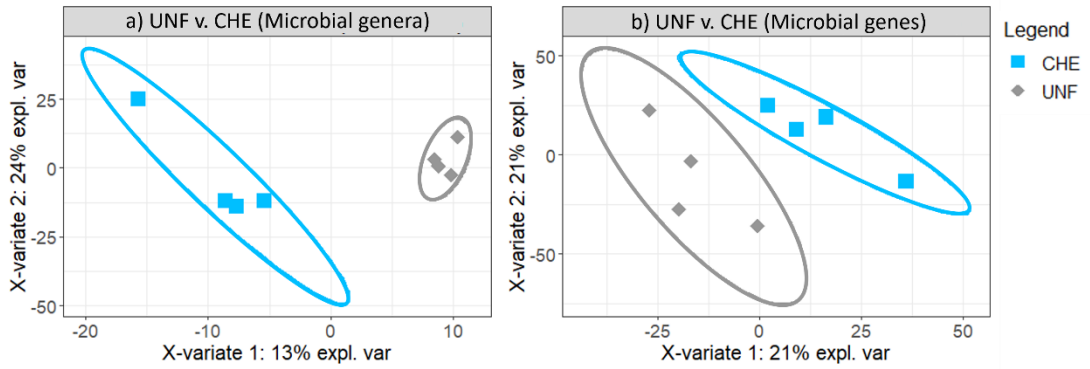


Figure 3. *Partial least squares discriminant analyses (PLS-DA) plots of individuals discriminating between uninfected and infected animals with high cumulative faecal egg count (UNF and CHE, respectively), using PLS-DA based on a) 899 microbial genera and b) 3124 microbial genes (i.e., KEGG level).*

The pairwise comparison of VAC with UNF in the PLS-DA analyses revealed that these groups are more similar than the corresponding comparisons of infected groups (CHE, and CLE) with UNF (Figure 4 vs. Figures 2 and 3). This is reflected in the overlap of the 95% confidence ellipses in the PLS-DA plot using microbial genes of the VAC vs. UNF comparison (Figure 4b).

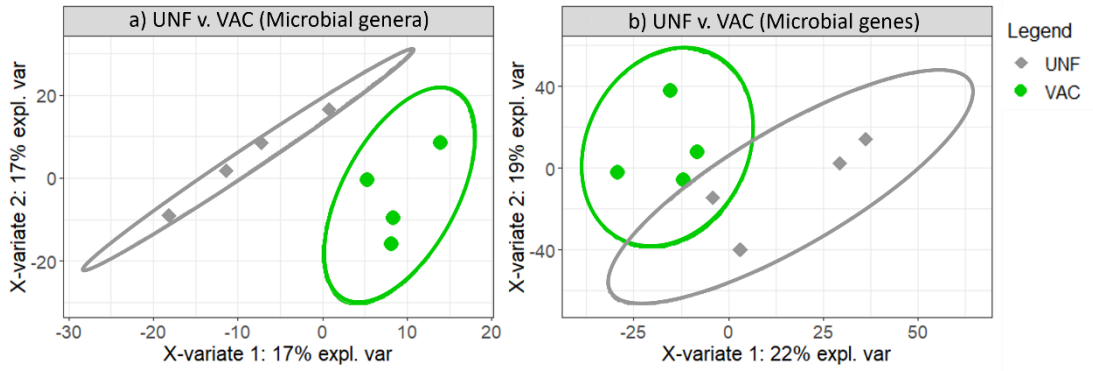


Figure 4. *Partial least squares discriminant analyses (PLS-DA) plots of individuals discriminating between uninfected and vaccinated animals (UNF and VAC, respectively), using PLS-DA based on a) 899 microbial genera and b) 3124 microbial genes (i.e., KEGG level).*

In comparing VAC with UNF, we observed increases of more than 40% in *Tenacibaculum*, *Thalassospira*, and *Pseudopropionibacterium* and decreases of 5%, 19%, and 0.5% in *Parvimonas*, *Macrococcus*, and *Arthrospira*, respectively, in VAC. For the same comparison, at the microbial genes-level, *TamB* (translocation and assembly module), *RTCA* (RNA 3'-terminal phosphate cyclase (ATP)), and *rhaA* (L-rhamnose isomerase) were 18%, 23%, and 15% more abundant, respectively, whereas *emrE* (small multidrug resistance pump), *modE*, and *nrfH* (cytochrome c nitrite reductase small subunit) were, respectively, 19%, 8%, and 39% less abundant in VAC.

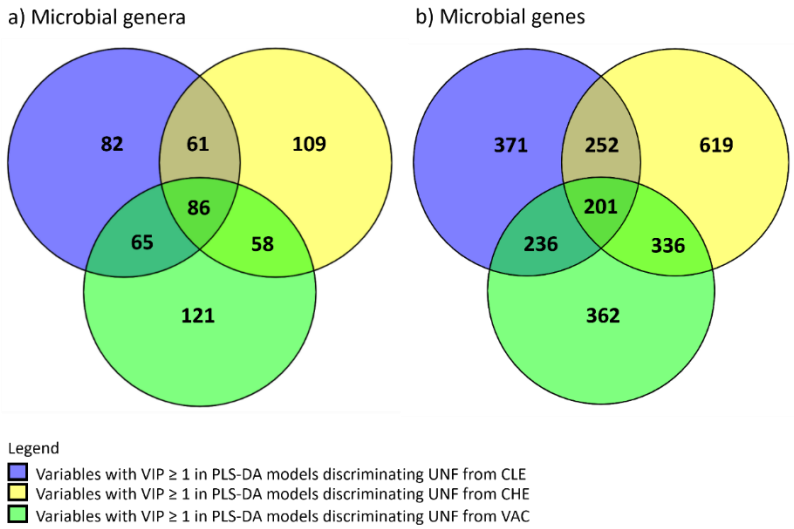


Figure 5. *Important variables for the discrimination of infected animals with low and high cumulative faecal egg count (CLE and CHE, respectively) and vaccinated (VAC) from uninfected (UNF) animals.*

A total of 86 microbial genera were important for the discriminations of CHE, CLE, and VAC from UNF (Figure 5), of which 36 were on average depleted and 38 were on average enriched in CHE, CLE, and VAC (of which the most abundant are presented in Figure 6).

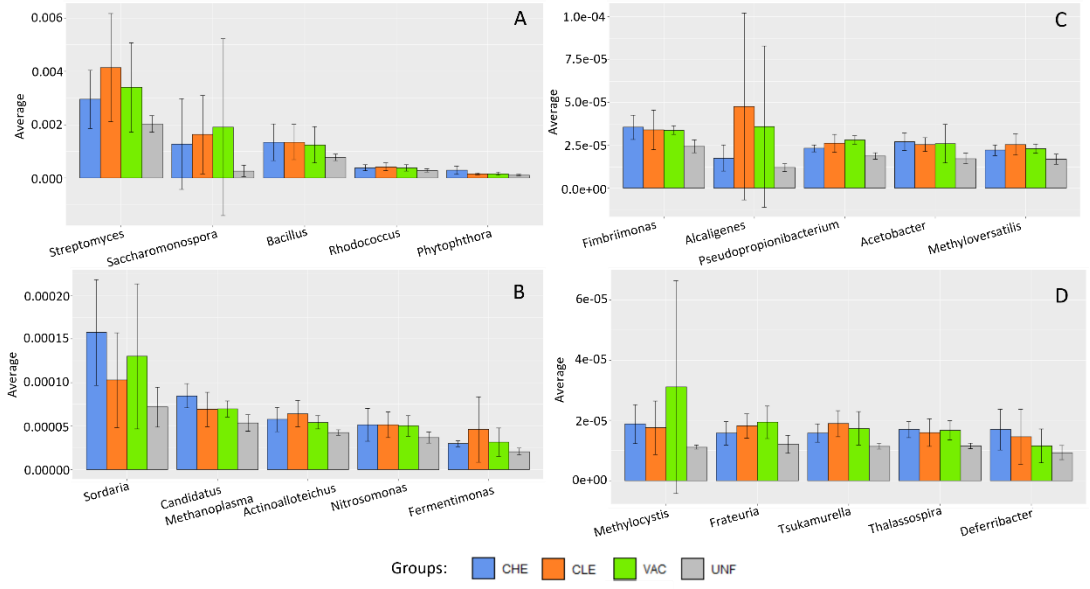


Figure 6. *Average relative abundance of the 20 microbial genera with highest relative abundances in infected showing high and low cumulative faecal egg count and vaccinated*

(CHE, CLE, and VAC), in comparison to uninfected (UNF) animals. Microbial genera were distributed from highest to lowest relative abundances in plots A to D, for clarity.

From 3124 microbial genes, 201 were identified in the PLS-DA models as important (VIP ≥ 1) to discriminate CHE, CLE, or VAC from UNF; 44 of these genes were depleted whereas 125 genes were enriched in CHE, CLE, and VAC, in comparison to UNF. The group of microbial genes depleted in the 3 groups in comparison to UNF included microbial genes *mtd*, *frhD*, *hdrA2*, *hxlB*, *fwdG*, *cofC* and *aksE* (i.e., methylenetetrahydromethanopterin dehydrogenase, coenzyme F420 hydrogenase subunit delta, heterodisulfide reductase subunit A2, 6-phospho-3-hexuloisomerase, 4Fe-4S ferredoxin, 2-phospho-L-lactate/phosphoenolpyruvate guanylyltransferase and methanogen homoaconitase small subunit, respectively), which belong to the methane metabolism pathway; microbial genes *hxlB*, *fwdG*, *mtd* and *hdrA2*, together with *gap2* (glyceraldehyde-3-phosphate dehydrogenase (NAD(P))), also participate in the carbon metabolism pathway; *AK6*, *nadX*, *MET8*, *pyrI* (i.e., adenylate kinase, aspartate dehydrogenase, precorrin-2 dehydrogenase / sirohydrochlorin ferrochelatase, and aspartate carbamoyltransferase regulatory subunit, respectively) belong to the biosynthesis of cofactors pathway (with genes *cofC* and *aksE*); *hxlB*, *aksE* and *gap2* are also part of the biosynthesis of amino acids pathway.

The group of microbial genes found to be enriched in CHE, CLE, and VAC in comparison to UNF included genes from several different pathways; for example, pathways associated with biosynthesis and metabolism of amino acids, such as biosynthesis of amino acids (*CTH*, *cysM*, *argA*, i.e., cystathionine gamma-lyase, S-sulfo-L-cysteine synthase (O-acetyl-L-serine-dependent), amino-acid N-acetyltransferase, respectively, and *thrB*), Selenocompound metabolism (*sat*, *cysNC,* i.e., sulfate adenylyltransferase and bifunctional enzyme CysN/CysC, respectively, and *CTH*); Glutathione metabolism (*pxpA*, *CARP*, i.e., 5-oxoprolinase (ATP-hydrolysing) subunit A and leucyl aminopeptidase, respectively); Cysteine and methionine

metabolism (*CTH*, *cysM*, and *speD*, i.e., S-adenosylmethionine decarboxylase); arginine biosynthesis (*argA*, and *GDH2*, i.e., glutamate dehydrogenase), and glycine serine and threonine metabolism (*CTH* and *thrB*). Some microbial genes in this group were associated with pathways of carbohydrate metabolism, such as the pyruvate metabolism (*bccA* and *nifV*, i.e., acetyl-CoA/propionyl-CoA carboxylase, biotin carboxylase, biotin carboxyl carrier protein, homocitrate synthase NifV); the pentose and glucuronate interconversions (*rhaD* and *rhaB*, i.e., rhamnulose-1-phosphate aldolase and rhamnulokinase, respectively); the fructose and mannose metabolism (*algA*, i.e., mannose-1-phosphate guanylyltransferase / mannose-6-phosphate isomerase, *rhaD*, *rhaA* and *rhaB*), and amino sugar and nucleotide sugar metabolism (*arnB*, *wbpA* and *glmU*, i.e., UDP-4-amino-4-deoxy-L-arabinose-oxoglutarate aminotransferase, UDP-N-acetyl-D-glucosamine dehydrogenase and bifunctional UDP-N-acetylglucosamine pyrophosphorylase / glucosamine-1-phosphate N-acetyltransferase, respectively, and *algA*).

Some microbial genes were associated with lipid metabolism, and belonged to pathways glycerophospholipid metabolism (*aas* and *glpQ*, i.e., acyl-[acyl-carrier-protein]-phospholipid O-acyltransferase / long-chain-fatty-acid--[acyl-carrier-protein] ligase and glycerophosphoryl diester phosphodiesterase, respectively), and fatty acid biosynthesis and metabolism (*fabI*, i.e., enoyl-[acyl-carrier protein] reductase I and *bccA*). Furthermore, some microbial genes were associated with biosynthesis of cofactors and vitamins, such as *pabA*, *queD*, *bioD*, *bioB*, *lipA*, *pncA*, *gltX*, *NQO1* (i.e., para-aminobenzoate synthetase component II, 6-pyruvoyltetrahydropterin/6-carboxytetrahydropterin synthase, dethiobiotin synthetase, biotin synthase, lipoyl synthase, nicotinamidase/pyrazinamidase, nondiscriminating glutamyl-tRNA synthetase and NAD(P)H dehydrogenase (quinone), respectively) and *fabI*, in the biosynthesis of cofactors pathway; *fabI*, *bioD* and *bioB*, also participating in the biotin metabolism pathway. Some genes had functions associated with energy sourcing such as *ndh* and *ATPVA* (i.e., NADH:ubiquinone reductase (H+-translocating) and V/A-type H+/Na+-

transporting ATPase subunit A, respectively) in the oxidative phosphorylation pathway, *GDH2* and *nirB* (i.e., nitrite reductase (NADH) large subunit), in the nitrogen metabolism pathway, and *sat* and *cysNC* (i.e., bifunctional enzyme CysN/CysC), in the sulfur metabolism pathway.

Microbial genes belonging to pathways associated with bacterial defence mechanisms, such as the monobactam biosynthesis (*sat* and *cysNC*), biofilm formation - *Vibrio cholerae* (*gspF*, *gspD*, and *rpoS*, i.e., general secretion pathway protein F, general secretion pathway protein D, and RNA polymerase nonessential primary-like sigma factor, respectively), biofilm formation - *Escherichia coli* and quorum sensing (*gspF* and *tatA* i.e., sec-independent protein translocase protein *TatA*), cationic antimicrobial peptide (CAMP) resistance (*arnT* and *amiABC*, i.e., 4-amino-4-deoxy-L-arabinose transferase and N-acetylmuramoyl-L-alanine amidase, respectively, and *arnB*), O-antigen nucleotide sugar biosynthesis (*algA*, *wbpA* and *glmU*), and lipopolysaccharide biosynthesis (*waaC*, *lpxJ* and *lpxC* i.e., heptosyltransferase I, Kdo2-lipid IVA 3' secondary acyltransferase and UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase, respectively, and *arnT*), together with microbial genes involved in environmental information processing, such as *gspD*, *rpoS*, and *oxyR*, (i.e., LysR family transcriptional regulator, hydrogen peroxide-inducible genes activator), belonging to the two-component system; *gspF*, *tatA*, and *gspD*, in the bacterial secretion system pathways and ABC transporters *gadC*, K14645 (i.e., glutamate:GABA antiporter and serine protease, respectively), and *arnB*, were also enriched in CHE, CLE, and VAC, in comparison to UNF.

5.4.2 Microbial genera and genes influenced by the presence of *O. ostertagi*

The variables identified as important for the discriminations of CLE and CHE from UNF were analysed together in a Venn diagram (Figure 7). Some microbiome features (147 microbial genera and 395 microbial genes) were important to discriminate UNF from infected animals independently of the level of resilience of the animals to the infection, however, most of them were

exclusively identified as important for the comparison CLE vs. UNF (167 microbial genera, and 939 microbial genes) or CHE vs. UNF (147 microbial genera and 665 microbial genes).

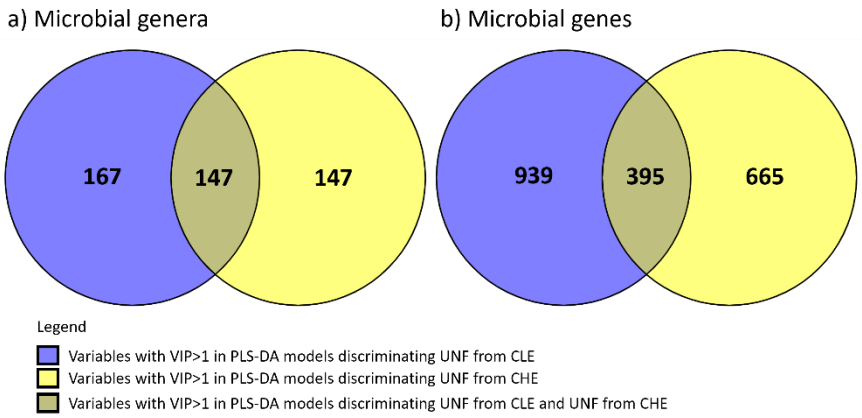


Figure 7. *Important a) microbial genera and b) microbial genes for the discrimination of infected animals with low and high cumulative faecal egg count (CLE and CHE, respectively) from uninfected animals (UNF).*

From a total of 899 genera, 167 and 147 were identified exclusively as important for the CLE vs. UNF and CHE vs. UNF discriminations, respectively (Figure 7a), suggesting a differential impact of the parasitism by *O. ostertagi*, depending on the resilience of the animal to the parasite. The group of microbial genera with VIP ≥ 1 in the CLE vs. UNF comparison and VIP < 0.5 in the CHE vs. UNF comparison included 67 microbes, all of them depleted in the CLE group, in comparison to UNF. Most of these microbes belonged to phyla Proteobacteria, e.g., *Aminobacter* and *Helicobacter*, Actinobacteria, e.g., *Micrococcus* and *Brevibacterium*, and Firmicutes, e.g., *Staphylococcus* and *Clostridium*. The group of microbes with VIP ≥ 1 in the CHE vs. UNF comparison (and VIP < 0.5 in the CLE vs. UNF comparison) included 63 microbial genera, of which 34 and 29 were enriched and depleted in CHE, respectively, in comparison to UNF. The 34 microbial genera included mostly Proteobacteria, e.g., *Desulfobacter* and *Sulfurospirillum*, Ascomycota, e.g., *Sporothrix*, and Firmicutes, e.g., *Sharpea*, whereas the 29 included mostly

Proteobacteria, e.g., *Methylococcus*, Firmicutes, e.g., *Dorea*, and Basidiomycota, e.g., *Cryptococcus*.

5.4.3 The rumen microbiome of infected differs from that of vaccinated animals

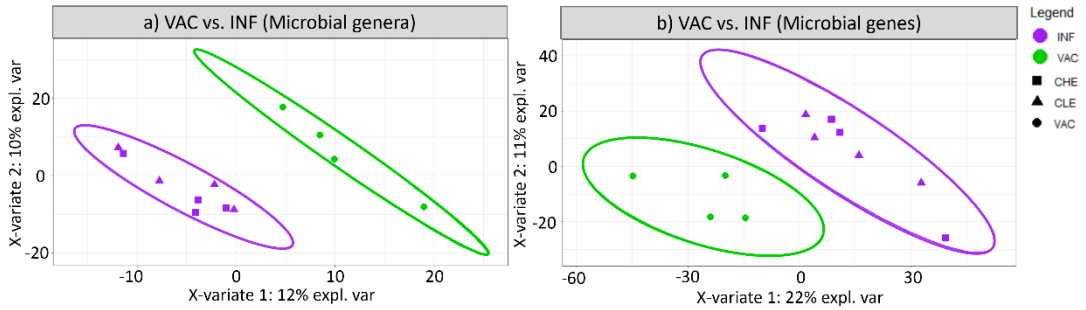


Figure 8. *Partial least squares discriminant analyses (PLS-DA) plots of individuals discriminating between vaccinated (VAC) and infected animals (INF, including animals with high and low cumulative faecal egg count, CHE, and CLE, respectively), using PLS-DA based on a) 899 microbial genera and b) 3124 microbial genes (i.e., KEGG level).*

The comparison of VAC with all infected animals (INF, combining CLE and CHE) resulted in a significant discrimination between these groups (Figure 8). A total of 344 microbial genera had a VIP ≥ 1 in the PLS-DA comparing VAC with INF (i.e., CHE and CLE). Within the 344, 176 microbial genera were enriched in INF, of which most belonged to phyla Proteobacteria (e.g., *Candidatus Phaeomarinobacter*, *Ruminobacter*, *Bartonella*, *Bosea*, *Advenella*, *Psychromonas*, and *Tateyamaria*, at least 20% more abundant on average in INF than in VAC), Actinobacteria (e.g., *Brevibacterium*, *Nocardiopsis*, *Brachybacterium*, and *Janibacter* showed at least 45% more relatively abundant in INF than in VAC), Bacteroidetes (e.g., INF had 45% or higher increases in the relative abundances of *Cytophaga*, *Emticicia*, *Leadbetterella*, *Petrimonas*, and *Solitalea*, in comparison to VAC), and Firmicutes (e.g., *Staphylococcus*, *Selenomonas*, and *Vagococcus* were 134%, 40% and 25%, respectively, more abundant in INF than in VAC), whereas 168 were enriched in VAC, most belonging to phyla Proteobacteria (e.g., *Moraxella*, *Acinetobacter*, and *Psychrobacter* were at least 75% more abundant in VAC than in infected animals), Ascomycota (e.g., *Verticillium*, and *Botrytis* were at

least 60% more abundant in VAC than in INF), Firmicutes (e.g., *Tepidanaerobacter*, *Carboxydothermus*, *Jeotgalibaca*, *Succiniclasticum*, *Pelosinus*, *Geosporobacter*, and *Acidaminococcus* were at least 20% more abundant in VAC than in INF), and Basidiomycota (e.g., *Trichosporon*, and *Moniliophthora*, 75% more abundant in VAC than in INF).

The analyses of this data at the functional level revealed 396 microbial genes important for the discrimination of VAC from INF, of which 246 were enriched in INF, whereas 150 were enriched in VAC. The microbial genes enriched in INF were associated with methane emissions (e.g., *frhD*, *mcrC*, and *mcrD*, i.e., coenzyme F420 hydrogenase subunit delta, methyl-coenzyme M reductase subunits C and D, respectively), carbon metabolism (e.g., *rbcL*, *fadN*, and *croR*, i.e., ribulose-bisphosphate carboxylase large chain, 3-hydroxyacyl-CoA dehydrogenase, and 3-hydroxybutyryl-CoA dehydratase, respectively) and biosynthesis of cofactors (e.g., *cbiT*, *ribB*, and *nadX*, i.e., cobalt-precorrin-6B (C15)-methyltransferase, 3,4-dihydroxy 2-butanone 4-phosphate synthase, and aspartate dehydrogenase, respectively), whereas the microbial genes enriched in VAC were mostly ABC transporters (e.g., *cbiN*, *cysW*, and *ccmA*, i.e., cobalt/nickel transport protein, sulfate/thiosulfate transport system permease protein, and heme exporter protein A, respectively), involved in quorum sensing (e.g., ABC.SP.S, *hfq*, and *lsrF*, i.e., putative spermidine/putrescine transport system substrate-binding protein, host factor-I protein, 3-hydroxy-5-phosphonooxypentane-2,4-dione thiolase, respectively), or included in the two-component system (e.g., *pilR*, *hyaC*, and *citX*, i.e., two-component system, NtrC family, response regulator PilR, Ni/Fe-hydrogenase 1 B-type cytochrome subunit, and holo-ACP synthase, respectively).

5.4.4 Co-abundance network of microbial genes in the rumen

The CLR-transformed abundances of 3124 microbial genes were investigated in a co-abundance network based on UNF records (Figure 9), in which clusters were identified that were found to be significantly enriched in microbial genes important (VIP ≥ 1) for the pairwise discriminations (i.e., PLS-DA models

comparing CHE, CLE or VAC, with UNF). Additionally, clusters were analysed for enrichment in microbial genes according to the biochemical pathways to which they belong.

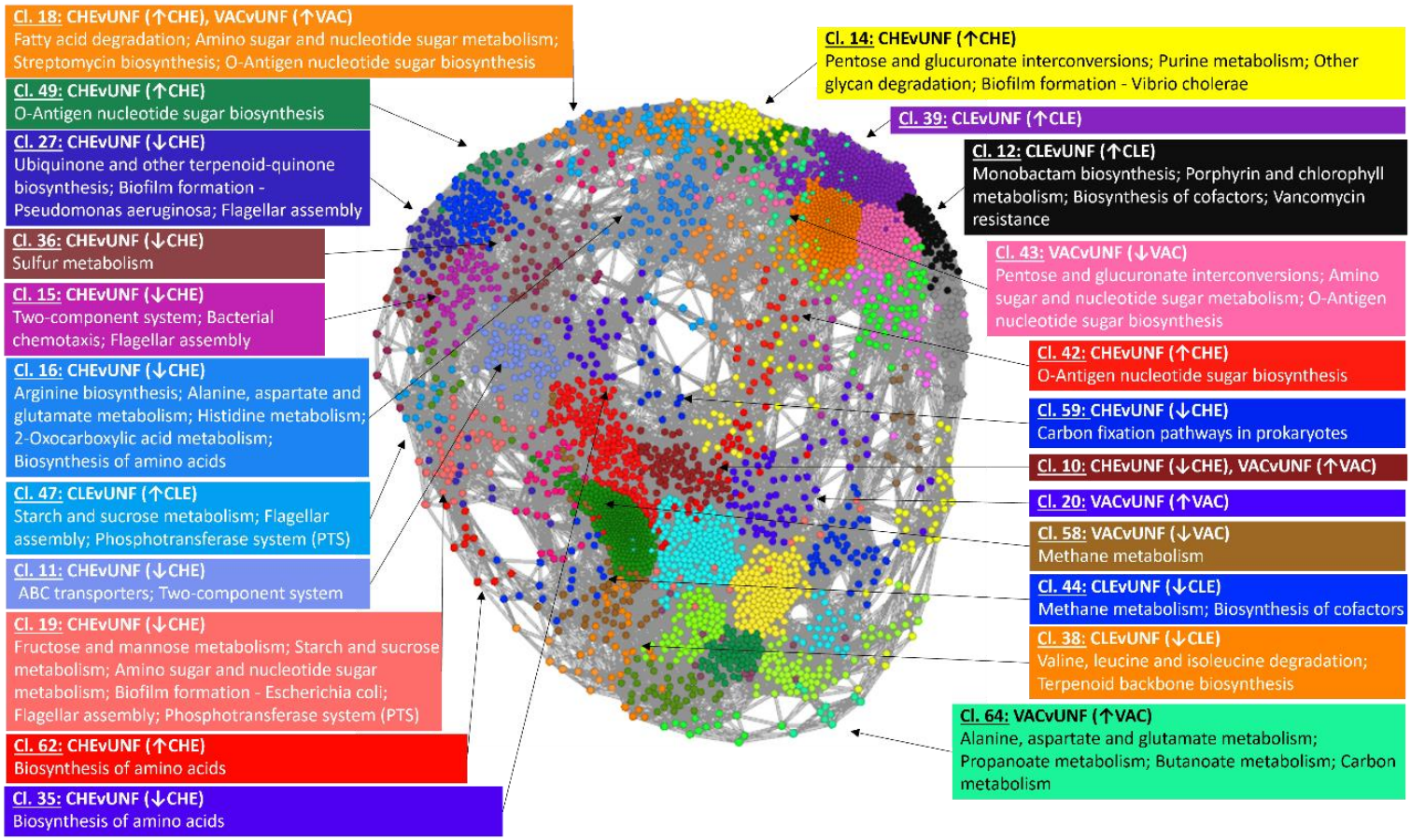


Figure 9. *Co-abundance network of microbial genes in uninfected animals (UNF). Each node represents the CLR-transformed abundance of a microbial gene, and each edge represents a correlation between nodes of r=0.98 or higher. Clusters significantly enriched in microbial genes identified as important (VIP ≥ 1) in the PLS-DA analyses for the pairwise discrimination of CHE, CLE, and VAC from UNF animals are labelled accordingly, and the increased or decreased abundance of microbial genes in each treatment group in comparison to UNF is indicated between the brackets by a ↑ or a ↓, respectively. Labels include information about enrichment regarding the biochemical pathways to which these microbial genes belong.*

The co-abundance network showed that clusters 12, 38, 39, 44, and 47 were significantly enriched in microbial genes important for the CLE vs. UNF discrimination, e.g., cluster 12 was enriched in microbial genes with increased abundance in CLE in comparison to UNF, and in microbial genes associated with monobactam biosynthesis; clusters 10, 11, 14, 15, 16, 18, 19, 27, 35, 36, 42, 49, 59, and 62 were significantly enriched in microbial genes important for the CHE vs. UNF discrimination, e.g., cluster 14 was enriched in microbial genes that, in comparison to UNF, had higher abundance in CHE, and in microbial genes associated with biofilm formation – *Vibrio cholerae*; clusters 20, 43, 58, and 64 were enriched in microbial genes important for the VAC vs. UNF discrimination, for example, cluster 64 showed enrichment of microbial genes with higher abundance in VAC than in UNF, and of microbial genes associated with Butanoate metabolism.

5.4.5 Methane

We observed enrichment of several methanogens in CHE and CLE, in comparison to UNF; *Candidatus Methanoplasma* had VIP ≥ 1 in both CHE vs. UNF and CLE vs. UNF comparisons. Other methanogens (e.g., *Methanococcus*, *Methanocorpusculum*, *Methanothermococcus*) were enriched in both CHE and CLE but at lower VIPs (0.8, 0.5, and 0.4 in the CHE vs. UNF and 0.5, 0.3, and 0.6 in the CLE vs. UNF comparisons, respectively). Additionally, methanotrophs *Methylocystis*, *Methylobacterium*, and *Methyloversatilis* were enriched (with VIP ≥ 1) in both CLE and CHE, in comparison to UNF, and other methanotrophs including *Methylomonas*, *Methylococcus*, and *Methylomicrobium* showed the same trends, although at lower VIPs (VIPs of 0.5, 1.0, and 0.8 in CHE vs. UNF, and 0.4, 0.5, and 0.3 in CHE vs. UNF comparisons, respectively). However, we also found depletion of methanogens, e.g., *Methanobacterium*, *Methanobrevibacter*, and *Methanosphaera* with VIP ≥ 1 in CHE and CLE, and *Methanomicrobium* (with VIPs of 0.8 and 1.6 in CHE vs. UNF and CLE vs. UNF comparisons, respectively).

Additionally, microbial genes associated with the methane metabolism pathway, e.g., *hxlB*, *comA*, *cofC*, and *aksE* (i.e., 6-phospho-3-hexuloisomerase, phosphosulfolactate synthase, 2-phospho-L-lactate/phosphoenolpyruvate guanylyltransferase, and methanogen homoaconitase small subunit, respectively), in cluster 44, and *mtrC,* and *mcrD* (i.e., tetrahydromethanopterin S-methyltransferase subunit C, and methyl-coenzyme M reductase subunit D, respectively), in cluster 56, *mcrA*, *mcrB* and *mcrC* (i.e., methyl-coenzyme M reductase alpha-, beta- and gamma-subunits, in cluster 1), important for the last step of methanogenesis, were depleted in CHE and CLE, in comparison to UNF (Figure 10).

Overall, some microbial genera and genes known to be associated with methane production were altered in infected in comparison to uninfected animals, however, we cannot, based on our analyses, conclude about the impact of these alterations on the methane production level of the host animals.

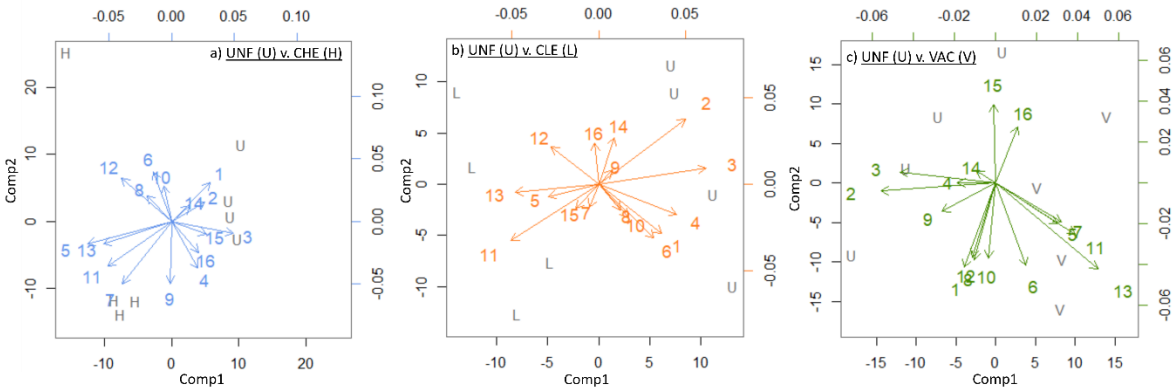


Figure 10. *Contribution of important methanogens and methanotrophs for the discrimination between uninfected (UNF) and a) infected showing high (CHE), and b) low (CLE) cumulative faecal egg count, and c) vaccinated and challenged (VAC). 1. Methanobacterium; 2. Methanobrevibacter; 3. Methanosphaera; 4. Methanomicrobium; 5. Candidatus Methanoplasma; 6. Methanosarcina; 7. Methanoculleus; 8. Methanococcus; 9. Methanocorpusculum; 10. Methanothermococcus; 11. Methylocystis; 12. Methylobacterium; 13. Methyloversatilis; 14. Methylomonas; 15. Methylococcus; 15. Methylomicrobium*

5.5 Discussion

5.5.1 Parasitism by, and vaccination against, the abomasal nematode *O. ostertagi* affects the rumen microbiota

This study is the first to focus on the changes to the rumen microbiome profiles both at taxonomic and microbial genes levels caused by the presence of the abomasal nematode *O. ostertagi* and/or by the vaccination against the nematode in dairy cattle.

The comparisons of treatment groups infected with high or low cumulative faecal egg counts (CHE, and CLE, respectively), and vaccinated (VAC) with uninfected (UNF) animals revealed 294, 314, and 330 microbial genera important for the CHE vs. UNF, CLE vs. UNF, and VAC vs. UNF discriminations, respectively.

A set of 86 microbial genera were identified as important (VIP ≥ 1) in the three pairwise comparisons (CHE vs. UNF, CLE vs. UNF, and VAC vs. UNF, Figure 5). Within this set, 38 microbial genera were enriched in CHE, CLE, and VAC. Within the 20 most abundant microbial genera enriched in CHE, CLE, and VAC, in comparison to UNF, we found microbial genera that were previously identified to be in association with gastrointestinal nematodes, such as *Bacillus* and *Deferribacter* (Figure 6). Although *Bacillus* spp. are known to produce fibrolytic enzymes that increase diet digestibility (Castillo-González et al., 2014), some strains of *B. thuringiensis* have been shown to have larvicidal activity against *Haemonchus contortus* (a nematode helminth parasitic of ruminants), to lead to a decrease in the relative abundance of *Fibrobacter succinigenes*, an important fibrolytic microbial genus in the rumen, and to impair diet degradability (Campos et al., 2019). Additionally, *Deferribacter* has previously been reported as enriched in the proximal colon microbiota of pigs infected with gastrointestinal nematode *Trichuris suis* (Li et al., 2012). Other microbial genera here enriched in CHE, CLE, and VAC, in comparison to UNF, were previously associated with digestive processes and feed conversion efficiency, e.g., *Acetobacter* is an acetogen previously found to be more abundant in cows producing milk with high concentration of saturated fatty

acids (SFA) in comparison to their low-SFA counterparts (Stergiadis et al., 2021), *Thalassospira* was previously reported as significantly increased in the caecum of high-nitrogen-utilizing goats (Wang et al., 2019), and *Streptomyces* is an actinobacteria able to produce ionophores, which are polyether carboxylic antibiotics that affect the cell membrane of Gram-positive bacteria, e.g., monensin (Castillo-González et al., 2014).

5.5.1.1 Microbial genera associated with efficient digestive functionality are depleted in the rumen of calves parasitised by *O. ostertagi*

Of the most important microbial genera for the discriminations of CLE and CHE from UNF (Figure 7), 70 were depleted in CHE and CLE, the majority belonging to phylum Firmicutes, e.g., *Enterococcus* and *Paenibacillus*, in class Bacilli; *Acidaminococcus* in class Negativicutes; and *Butyrivibrio*, *Eubacterium*, *Faecalicatena*, *Lachnospira* and *Pseudobutyrivibrio* in class Clostridia.

Enterococcus is a lactate-producing bacterium (Nagpal et al., 2015), belonging to the class Bacilli, phylum Firmicutes. Whereas some *Enterococcus* species are opportunistic pathogens of humans (e.g., *E. faecalis* and *E. faecium*) or animals (e.g., *E. hirae*; Lebreton, Willems, and Gilmore (2014)), with *Enterococcus* having been previously reported as enriched in the faecal microbiota of cats parasitised by the nematode *Toxocara cati* (Duarte et al. 2016), they are also considered normal commensals of the gastrointestinal tract, and for example, Jackson et al. (2011) reported that 88.7% of 718 dairy cattle faecal samples were positive for enterococci. Also in the class Bacilli, *Paenibacillus* is an hemicellulose-degrading bacterium, identified in the core rumen fluid microbiome (although at low relative abundances) of dairy cattle (Wirth et al., 2018).

Belonging to the class Negativicutes, *Acidaminococcus* is an amino acid degrader, here depleted in the rumen of infected animals. This microbial genus was previously shown to be important for the pyruvate to phosphoenolpyruvate conversion (gluconeogenesis) and shown to be associated with high methane emissions (Auffret et al., 2018) and low feed conversion efficiency (Auffret et al., 2020) in beef cattle.

Several microbial genera in the class Clostridia were also depleted in the rumen of animals parasitised by *O. ostertagi*. For example, *Butyrivibrio* and *Pseudobutyrivibrio* are commonly found in the rumen and can utilize xylans and pectins, and ferment carbohydrates into butyrate, formate, lactate and acetate, being important energy suppliers to ruminants. *Butyrivibrio* is also involved in protein breakdown and biohydrogenation of fatty acids (Palevich et al., 2020). *Butyrivibrio* and *Pseudobutyrivibrio* were also previously shown to have the largest number of *LuxS* protein genes in their genomes and have the ability to use the *Lux*-based AI-2 quorum sensing genes, which is the most abundant and predominant communication system used by rumen bacteria (Won et al., 2020). In the present study, however, although *Butyrivibrio* and *Pseudobutyrivibrio* were depleted, the microbial gene *luxS* (cluster 3) was enriched in infected groups. *Eubacterium* (class Clostridia), which was also depleted in inflected animals, has previously been associated with high feed conversion efficiency in beef cattle (Auffret et al., 2020). This genus includes cellulolytic, e.g., *Eubacterium cellulosolvens* (Prins et al., 1972), and hemicellulolytic and xylanolytic groups e.g., *Eubacterium ruminantium* (Taguchi et al., 2004), which are important lactate and butyrate producers (Flint et al., 2014) of the rumen core microbiome, and has been shown to have high metabolic versatility, being involved e.g., in the carbohydrate metabolic process, gluconeogenesis, and glycolysis (Wirth et al., 2018). Another microbial genus in the class Clostridia found here to be depleted in infected animals is *Lachnospira*. This microbial genus belongs to the family *Lachnospiraceae*, which has previously been identified as a core member of the rumen microbiome of beef cattle (Li and Guan, 2017). This microbial genus was also previously reported as relatively more abundant in low feed conversion efficiency animals and as associated with pathways such as methane metabolism, glyoxylate and dicarboxylate metabolism, tryptophan metabolism and valine leucine and isoleucine degradation. Furthermore, the family *Lachnospiraceae* was previously reported as increased in microbiota samples from colorectum bile ducts of hamsters parasitised by the nematode *Opisthorchis viverrini* (Plieskatt et al., 2013).

The results suggest that the parasitism by *O. ostertagi* causes changes on the rumen microbiome of the host animal, leading to the depletion of several community members strongly associated with fermentation and digestive-associated biochemical pathways, which could potentially be unfavourable for the animal.

5.5.1.2 The rumen microbiota of *O. ostertagi*-parasitised calves is enriched in microbial genera with pathogenic potential

The microbial genera enriched in CHE and CLE in comparison to UNF were in their majority from the phylum Actinobacteria, i.e., bacteria with Gram-positive or Gram-positive like walls with pathogenic potential. For example, microbial genera *Streptomyces*, *Tsukamurella*, *Pseudopropionibacterium* and *Segniliparus* include human pathogenic species such as *S. somaliensis* (Kirby et al., 2012), *T. paurometabolum* (Shapiro et al., 1992), *P. propionicum* (Suzuki et al., 2019), *S. rotundus* and *S. rugosus* (Kim et al., 2013). Other members of the phylum Actinobacteria were also enriched in infected animals. For example, *Curtobacterium*, which includes plant pathogen species reported as causing infection in humans (Francis et al., 2011), *Kocuria* is a potential human pathogen (Kandi et al., 2016), and *Dermabacter* is a rare human pathogen (Gómez-Garcés et al., 2001). Some microbial genera (also Actinobacteria) potentially cause dysbiosis in the rumen, as they were here enriched in infected animals, e.g., the genus *Kutzneria* is represented by *K. albida* which produces the secondary product aculeximycin, an antibiotic with activity against Gram-positive bacteria and fungi (Rebets et al., 2014), and the genus *Saccharothrix* produces a wide variety of potent antibiotics with activity against bacteria and yeasts (Strobel et al., 2012). Both *Kutzneria* and *Saccharothrix* were recently isolated from bovine rumen samples, and found to be more abundant in dairy cattle with high saturated fatty acid (SFA) content in their milk, in comparison to their low-SFA counterparts (Stergiadis et al., 2021). Additionally, *Saccharomonospora* (phylum Actinobacteria) was previously reported as associated with increased methane emissions in beef cattle (Martínez-Álvaro et al., 2020).

The enrichment of potentially pathogenic microbial genera in the rumen of CHE and CLE in comparison to UNF suggests that the presence of the parasite has led the rumen microbiome into a state of dysbiosis, which could, in addition to the direct effect of *O. ostertagi* on animal health and welfare, contribute to the reduced productivity previously observed in infected animals.

5.5.1.3 Rumen microbiota affected by the vaccine against *O. ostertagi*

We identified 31 microbial genera which were important for the discrimination of VAC from UNF but not important (VIP < 0.5) for the discrimination of UNF from CHE or CLE. Of these, 9 and 22 microbial genera were depleted and enriched, respectively, in VAC. The microbial genera depleted in VAC belonged mostly to the Bacteroidetes phylum, e.g., *Barnesiella*, *Odoribacter*, and *Parabacteroides*. *Barnesiella* was previously identified in the rumen (Islam et al., 2021), *Odoribacter* is highly heritable and strongly positively genetically correlated with methane production (Martínez-Álvaro et al., 2021); *Parabacteroides* has been shown to be depleted in the faecal microbiome of mice throughout the course of infection by the nematode *Trichuris muris*, and the authors linked this change with the reduced plant-derived carbohydrates metabolism, and immune and signalling responses that may have led to increased amino acid content in stool samples (Houlden et al., 2015).

Microbial genera enriched in VAC, in comparison to UNF, were mostly from phyla Ascomycota and Basidiomycota. Ascomycota phyla included, e.g., *Colletotrichum*, *Botrytis*, *Verticillium*, and *Thielavia*; *Colletotrichum* is widely recognized genus of phytopathogenic fungi (da Silva et al., 2020), however, some species have previously been reported as causative agent of disease in humans (Cano et al., 2004); *Botrytis* is a necrotrophic phytopathogen (Grant-Downton et al., 2014). *Veticillium* is an endoparasite of nematodes (Barron, 1987; Segers et al., 1994). *Thielavia* has been isolated from a case of keratitis (Theoulakis et al., 2009) and as a cause of fatal brain infection in an Indian farmer (Badali et al., 2011). Basidiomycota phylum included e.g., *Tsuchiyaea*, which was previously shown to have a positive association with methane emissions (Martínez-Álvaro et al., 2020), and *Punctularia* and

Cutaneotrichosporon, recently reported as enriched in the rumen microbiome of cattle with in high SFA content in comparison to their low-SFA counterparts (Stergiadis et al., 2021).

We evaluated the impact of the vaccine by comparing the microbiome profiles of vaccinated with those of infected animals (INF = CHE + CLE), to provide an insight of the possible outcomes of vaccination on the rumen microbiome under the condition that all animals were infected with *O. ostertagi* at the same dose.

The analyses of the rumen microbiome profiles at the taxonomic level revealed several microbial genera to be increased in INF, in comparison to VAC, as for example, the potential pathogen *Bartonella* (Schmidt and Hensel, 2004), and *Bosea*. The latter genus was recently reported to be associated with a nosocomial infection (Skipper et al., 2020). Both *Bartonella* and *Bosea* belong to the phylum Proteobacteria, previously shown to be enriched in the rumen of low-nitrogen-utilizing goats (i.e., with higher nitrogen waste in the form of urea (Wang et al., 2019)). Nitrogen retention has previously been reported to be decreased in parasitised calves, mostly due to increased excretion through urine and faeces (Parkins et al., 1982), and Fox et al. (1989a) reported that the nitrogen digestibility of infected calves was significantly lower than that of ad-libitum-fed and pair-fed calves and attributed this to changes in the protein metabolism of infected animals. Furthermore, microbial genes involved in the nitrogen metabolism pathway, such as *nifD* and *nifK* (nitrogenase molybdenum-iron protein alpha and beta chains, respectively), were depleted in both CHE and CLE, in comparison to UNF. Several microbial genera enriched in the rumen of INF in comparison to VAC belonged to the phylum Actinobacteria, e.g., *Brevibacterium*, which was also enriched in CHE and CLE, in comparison to UNF.

Regarding the microbiome functional profiles, the results suggested that vaccinating the calves against *O. ostertagi* could potentially lead to lower methane emissions, as well as strengthen the environmental sensing potential

and bacterial cell wall structures-associated biochemical pathways, important for the healthy functioning of the rumen microbial communities. Enriched presence of microbial genes associated with environmental sensing and bacterial cell wall structures was also observed in UNF, in comparison to CHE and CLE. These results suggested that the vaccine prevented some of the alterations observed in the unvaccinated infected groups CHE and CLE and underlined that VAC-derived microbiome profiles are more like those of uninfected than those derived from CHE or CLE.

5.5.2 Rumen microbiome functionality is affected by *O. ostertagi* parasitism, and by the vaccine against the abomasal nematode

We compared the microbial genes derived from the rumen of CHE, CLE, and VAC with those from UNF animals, and found that the presence of the parasite not only alters the microbial community profiles in the rumen, but also their functional potential, particularly affecting digestive processes such as those associated with amino acid degradation, and putatively protein deposition, fatty acids processing, and carbohydrate uptake regulation, bacterial cell wall structures and environmental sensing, and even methane emissions.

5.5.2.1 Microbiome-associated digestive processes in the rumen are affected by the abomasal nematode

One of the most common clinical signs of nematode parasitism in ruminants is the animals' inappetence and associated weight loss; Fox et al. (1989) used pair-fed animals to shown that the weight loss of parasitised animals is not exclusively associated with their inappetence, but also with gastrointestinal function impairment, particularly considering the increased abomasal pH due to damage to the gastric glands by the parasite, and the importance of the acidic pH for the activation of peptidases in the abomasum (with consequences on nitrogen digestibility).

Histidine is one of the limiting amino acid in ruminants, particularly in young cattle (Greenwood and Titgemeyer, 2000; Onodera, 2003), making bacterial histidine biosynthesis essential for ruminants, particularly in environments

where it is present at very low levels; McCuistion et al. (2004) showed increased nitrogen retention in cattle supplemented with histidine, suggesting histidine to be crucial for protein deposition. Microbial genes in the histidine metabolism pathway were enriched in cluster 16 (Figure 9), some of these belonging to the *his* operon (e.g., *hisH* and *hisIE*, imidazole glycerol-phosphate synthase subunit *HisH*, and phosphoribosyl-AMP cyclohydrolase / phosphoribosyl-ATP pyrophosphohydrolase, respectively), involved in the first step of the pathway (from PRPP to histidine), which in turn forms a critical link between amino acid, purine and thiamine biosynthesis (Ames et al., 1961; Winkler and Ramos-Montanez, 2009). These microbial genes, along with *hdc* and *urdA* (i.e., histidine decarboxylase, and urocanate reductase, respectively) were depleted in CHE (and CLE) animals.

Microbial genes involved in arginine biosynthesis were also included in cluster 16, e.g., *argC* and *argG* (i.e., N-acetyl-gamma-glutamyl-phosphate reductase, and argininosuccinate synthase, respectively), which were depleted in CHE and VAC, but enriched in CLE, and e.g., *argB* and *argH* (i.e., acetylglutamate kinase, and argininosuccinate lyase, respectively), which were depleted in CHE, CLE, and VAC. Arginine is a fundamental amino acid, with functions associated with the urea cycle, and protein synthesis; the post-ruminal supplementation of arginine has been shown to increase lactation performance and nitrogen utilization efficiency in lactating dairy cattle (Ding et al., 2019), and L-arginine in-vitro supplementation of rumen digesta collected from rumen-fistulated cows has been shown to contribute to rumen fermentation efficiency, leading to increased production of volatile fatty acids (Chacher et al., 2012).

Microbial genes *arG* and *argH* also participate in the alanine, aspartate, and glutamate metabolism, together with *gltB*, *carA*, and *carB* (i.e., glutamate synthase (NADPH) large chain, and carbamoyl-phosphate synthase small and large units, respectively, in cluster 16), which were depleted in CHE, but enriched in CLE. Other microbial genes in the same pathway were depleted in both CHE and CLE, e.g., *mhpE* and *AROA2* (i.e., 4-hydroxy 2-oxovalerate

aldolase, and 3-deoxy-7-phosphoheptulonate synthase, respectively), in cluster 35, *aroD* (i.e., 3-dehydroquinate dehydratase) in cluster 62, and *puuE* (i.e., 4-aminobutyrate aminotransferase), in cluster 64. Glutamate was previously shown to be decreased in the rumen of dairy cows with high feed intake, in comparison to their low feed intake counterparts (Li et al., 2020), and free glutamate in formula has been shown to increase satiation and satiety levels in human infants, leading to significantly lower intake (Ventura et al., 2012). The depletion of microbial genes associated with glutamate metabolism in infected animals potentially leads to increased glutamate levels in the rumen, thus being one of the potential contributors to the inappetence typically observed in cattle parasitised by *O. ostertagi*.

Valine, leucine, and isoleucine act as hydrogen donors in fermentations carried out by proteolytic anaerobes, and they are processed into fatty acids within the rumen, having an additive effect on cellulose digestion (Dehority et al., 1958; Menahan and Schultz, 1964). Additionally, these fatty acids have been shown to greatly stimulate the growth of many non-cellulolytic bacteria, and Allison et al., (1958) showed that they were a requirement for cellulolytic cocci *Ruminococcus flavefaciens* and *Ruminococcus albus*. In the present work, *Ruminococcus* was found to be on average 6.8% and 13.7% more abundant in CLE and CHE, respectively, in comparison to UNF. Additionally, cluster 38 was significantly enriched in microbial genes in the valine, leucine, and isoleucine degradation pathway, including genes E6.4.1.4B and *liuC* (3-methylcrotonyl-CoA carboxylase beta subunit and methylglutaconyl-CoA hydratase *liuC*, respectively), which were more abundant in CLE (and non-significantly in CHE) than in UNF.

Although some microbial genes with functions associated with biosynthesis of amino acids were more abundant in CLE animals, the majority was depleted in CHE and CLE, suggesting that protein metabolism processes may be hindered, in agreement with Fox et al. (1989a), who reported a 22% decrease in apparent nitrogen digestibility in infected, in comparison to pair-fed calves, accompanied by increased nitrogen excretion.

The PTS system catalyses the uptake and phosphorylation of many different carbohydrates and plays a major role in the carbon catabolite repression (CCR); for each organism, a specific hierarchy exists for the utilization of carbon sources; glucose, fructose or sucrose are usually the preferred sugars, and their availability in the environment represses the uptake and utilization of other carbon sources (Deutscher et al., 2006). Cluster 19 was enriched in microbial genes involved in PTS, due to the inclusion of *celA* and *treB* (i.e., cellobiose PTS system EIIB component, and trehalose PTS system EIIBC or EIIBCA component, respectively), also involved in the starch and sucrose metabolism pathway, and *manXa*, *manX*, *manY*, *manZ* (i.e., mannose PTS system EIIA, EIIAB, EIIC, and EIID components, respectively), *cmtA* and *cmtB* (i.e., mannitol PTA system EIIA, and EIICBA or EIICB components, respectively), also involved in the fructose and mannose metabolism pathway; whereas cluster 19 was enriched in microbial genes significantly depleted in CHE, 9 out of 10 microbial genes associated with the PTS in this cluster were more abundant in CHE and CLE than in UNF. In cluster 47 there was enrichment of microbial genes with significantly increased abundance in CLE, and of microbial genes in the PTS pathway, that were more abundant in infected animals than in uninfected (e.g., *agaF* and *celB*, respectively, N-acetylgalactosamine PTS system EIIA component, and cellobiose PTS system EIIC component). Considering that the network was constructed based on the abundances of microbial genes derived from UNF animals only, the results suggest that parasitism by *O. ostertagi* leads to alterations in microbial genes in clusters 19 and 47, particularly to an increase in the abundance of microbial genes in the PTS system.

5.5.2.2 Previously identified rumen appetite- and performance-associated microbial gene biomarkers are affected by the presence of the abomasal nematode

Microbial genes previously identified as feed conversion efficiency biomarkers were evaluated here for their association to the parasitism levels; *infA*, *galK*, and *xylE* (translation initiation factor IF-1, galactokinase, and MFS transporter, SP family, xylose:H+ symporter, respectively) were important for the CHE vs.

UNF and CLE vs. UNF discriminations, and enriched in CHE and CLE, and were previously shown to be associated with decreased feed conversion efficiency, whereas *dph2*, *hupB*, ABC.SS.P, *cbiN*, *fliN*, *aor*, *zntA*, *mcp* and *oadA* (diphthamide synthase subunit DPH2, DNA-binding protein HU-beta, simple sugar transport system permease protein, cobalt transport protein, flagellar motor switch protein FliN/ FliY, aldehyde:ferredoxin oxidoreductase, Cd2+/Zn2+-exporting ATPase, methyl-accepting chemotaxis protein, and oxaloacetate decarboxylase, alpha subunit, respectively), here depleted in infected animals, were previously shown to be associated with increased feed conversion efficiency (Lima et al., 2019). Additionally, previously identified biomarkers for increased appetite (*rfbG* and *rfbF*, i.e., CDP-glucose 4,6-dehydratase and glucose-1-phosphate cytidylyltransferase) and increased weight gain (*slyD*, i.e., FKBP-type peptidylprolyl cis-transisomerase *SlyD, hupB*, and *dph2*) were here depleted in infected animals, whereas biomarkers for decreased appetite (*rpoB*, INO1, and *rdgB*, i.e., DNA-directed RNA polymerase subunit beta, myo-inositol-1-phosphate synthase, and XTP/dITP diphosphohydrolase, respectively) and for decreased growth (*rpmJ*, and *atpH*, i.e., large subunit ribosomal protein L36, and F-type H+-transporting ATPase subunit delta, respectively, and amiABC) were enriched in *O. ostertagi* parasitised animals (Lima et al., 2019).

5.5.2.3 Ruminal bacterial cell wall structures and environmental information processing affected by *O. ostertagi* parasitism

Several ABC transporters were important for the discrimination of CHE and CLE from UNF, some of these were included in cluster 10, such as *livK*, *afuA*, *metQ* and *togM* (i.e., branched-chain amino acid transport system substrate-binding protein, iron (III) transport system substrate binding protein, D-methionine transport system substrate-binding protein, and oligogalacturonide transport system permease protein, respectively), which were depleted in CHE and CLE but enriched in VAC, in comparison to UNF, along with e.g., *tcyN* and *lplB* (i.e., L-cystine transport system ATP-binding protein, and putative aldouronate transport system permease protein, respectively), in cluster 11. Also in cluster 11, we found ABC transporters depleted in CHE, CLE, and VAC,

such as *gsiC* and *lysY* (i.e., glutathione transport system permease protein, and putative lysine transport system ATP-binding protein, respectively). Additionally, we found enrichment of microbial genes associated with the two-component system in cluster 11 and 15, most of which were depleted in CHE and CLE, e.g., *baeS*, *kdpD*, and *kdpE* (i.e., two-component system, OmpR family, sensor histidine kinase *BaeS* and *KdpD*, and two-component system, OmpR family, KDP operon response regulator *KdpE*, respectively).

Microbial genes in pathways associated with flagellar assembly were altered by the presence of the nematode and by the vaccination, for example, in cluster 19, microbial genes *flgK*, *flgL* and *fliN* (i.e., flagellar hook associated proteins 1 and 3, and a flagellar motor switch protein, respectively), depleted in infected animals, and in cluster 27, e.g., flagellar biosynthesis protein *FlhA*, *Flip*, and *FliQ*, enriched in infected and VAC animals. Interestingly, microbial genes associated with flagellar assembly in cluster 15, e.g., flagellar basal-body rod protein *flgB*, flagellum-specific ATP synthase *fliI*, and flagellar biosynthesis protein *fliR* were found to be enriched in CLE and VAC, but depleted in CHE, showing the differential effect of the nematode according to the level of parasitism.

Infected animals showed enrichment of microbial genes associated with bacterial defence mechanisms, e.g., microbial genes *rbfA* and *rbfC* (i.e., glucose-1-phosphate thymidylyltransferase, and dTDP-4-dehydrorhamnose 3,5-epimerase, respectively), associated with streptomycin resistance (cluster 18), *dapA* (i.e., 4-hydroxy-tetrahydrodipicolinate synthase, in cluster 12), associated with monobactam biosynthesis, *vanX* and *vanW* (i.e., zinc D-Ala-D-Ala dipeptidase, and a vancomycin resistance protein, respectively), associated with vancomycin resistance (cluster 12). Vancomycin is an antibiotic compound that acts on the cell walls, inhibiting the PG synthesis by interacting with the PG precursor N-acetylmuramyl-pentapeptide on the exterior surface of the cell wall. Resistance to vancomycin is most often found in Gram-positive enterococci, and it occurs by modification of the pentapeptide (replacement of a C terminal D-ala residue by a D-lactate or a D-serine) so

they have low affinity to vancomycin or by degradation of these precursors by production of an enzyme (Courvalin, 2006). Vancomycin resistance is acquired through horizontally transferred plasmid inducible genetic elements; antibiotic-resistant bacteria are an increasing public threat and the cattle rumen has been previously identified as a reservoir for microbial genes associated with antimicrobial resistance (AMR, Auffret et al. 2017). Auffret et al. (2017) suggested that the higher abundance of AMR genes observed in the rumen of concentrate-fed cattle was due to the change in diet from forage-fed, and that the presence of these genes was associated with the adaptation of microbial population to a recently altered environment. Although the nematode infection challenge differs substantially from that of altering the diet composition, our results are in line with this suggestion.

Microbial genes associated with O-antigen nucleotide sugar biosynthesis in cluster 18 (e.g., *wecB* and *gmd*, i.e., UDP-N-acetylglucosamine 2-epimerase (non-hydrolysing), and GDPmannose 4,6-dehydratase, respectively), in cluster 42 (e.g., *wbpB* and *wbpD*, i.e., UDP-N-acetyl-2-amino-2-deoxyglucuronate dehydrogenase, and UDP-2-acetamido-3-amino-2,3-dideoxy-glucuronate N-acetyltransferase, respectively), and in cluster 49 (e.g., *pseI*, i.e., pseudaminic acid synthase) were mostly enriched in CHE and CLE, and depleted in VAC, in comparison to UNF.

5.5.3 The influence of O*. ostertagi* on the rumen microbiome differs according to the degree of parasitism

Our research suggests that the parasitism of the abomasal nematode *O. ostertagi* in cattle had significant effect on rumen microbiome profiles, and that this impact differs according to the degree of parasitism. At the microbiota level, we observed that microbial genera exclusively important for the CHE vs. UNF discrimination enriched in CHE mostly belonged to phylum Proteobacteria, Ascomycota, and Firmicutes, whereas no microbial genus was identified as important exclusively for the CLE vs. UNF discrimination and enriched in CLE. On the other hand, microbial genera depleted in CHE (CHE vs. UNF discrimination) belonged to the phyla Proteobacteria, Firmicutes and

Basidiomycota, whereas those depleted in CLE (CLE vs. UNF discrimination) were mostly Proteobacteria, Actinobacteria, Bacteroidetes, and Firmicutes. These results suggest that the enrichment of some microbial communities only occurred in the rumen of animals less resistant to the infection challenge (CHE), and that the depletion of groups differed according to the level of parasitism. At the functional level, we found that microbial genes involved in amino acids metabolism, i.e., arginine biosynthesis, alanine, aspartate and glutamate metabolism, histidine metabolism, and microbial genes involved in environmental sensing such as those in pathways of two-component systems, bacterial chemotaxis, and ABC transporters, were depleted in CHE, whereas microbial genes in pathways such as fatty acid degradation, O-antigen nucleotide sugar biosynthesis, pentose and glucuronate interconversions, and purine metabolism were enriched in CHE. On the other hand, alterations associated with low levels of parasitism (i.e., CLE) involved microbial genes associated with the methane metabolism, biosynthesis of cofactors, and the valine, leucine and isoleucine degradation pathways (depleted in CLE), and monobactam biosynthesis and vancomycin resistance (enriched in CLE).

The CHE and CLE animals were selected from a wider group of calves, based on their extreme cFEC. This variation in cFEC is most likely due to the natural variation in the resistance of the animals to the parasite, associated, for example, with their innate immune systems' differential ability to respond to the parasite's presence, which could be the reason for the differences in the impact of the nematode on the rumen microbiome reported here. However, during its life cycle, *O. ostertagi* passes through the rumen, where L3 larvae undergo the exsheathment process, developing into L4 and exiting towards the abomasum; the efficiency of this process (i.e., % of L3 that successfully develop into L4 and leave the rumen) depends on the conditions found in the rumen (e.g., the pH and associated bicarbonate/carbonic acid buffering system act as stimuli, Derosa et al. 2005), and thus, the difference in cFEC between CHE and CLE could therefore be the result of a natural variation of

the rumen microbiome between more and less resistant animals that existed previously to the infection.

5.5.4 Methanogens and methanotrophs are influenced by *O. ostertagi* parasitism

Previous work showed decreased methane output in nematode parasitised ruminants, mostly due to decreased feed intake, whereas methane yield was increased in parasitised animals (Houdijk et al., 2017; Fox et al., 2018). Additionally, 17 peptides associated with the free-living phase of *O. ostertagi* were previously shown to influence methane metabolism (Heizer et al., 2013). Our results suggest that the increased methane yield observed in parasitised animals (even in comparison to restricted fed unparasitized controls) is at least partially associated with alterations in the rumen microbiota community structure caused by the presence of the nematode, as suggested previously by (Fox et al., 2018). The 20 most abundant microbial genera enriched in CHE, CLE, and VAC, in comparison to UNF, included taxa that have previously been associated with methane production. For example, *Saccharomonospora*, a glycopeptide-producing *Actinobacterium* (Donadio and Sosio, 2009) that was previously reported as positively associated with CH_4 emissions in beef cattle (Martínez-Álvaro et al., 2020), *Tsukamurella*, an *Actinobacterium* previously reported as negatively associated with CH_4 emissions in beef cattle (Auffret et al., 2018), *Methylocystis*, a facultative methylotroph (Haque et al., 2020) and *Nitrosomonas*, an epimurial bacterium that oxidizes methane and ammonia produced by ureolytic bacteria (Mitsumori et al., 2002), were here enriched, whereas *Methylomonas*, previously observed to negatively correlate to CH_4 in the rumen of beef cattle (Auffret et al., 2018), was depleted in CHE, CLE, and VAC, in comparison to UNF. Furthermore, the comparison of VAC with INF showed that INF had enrichment of microbial genes with functions associated with methane emissions.

5.6 Conclusions

Our results suggest that infection by the abomasal nematode *Ostertagia ostertagi*, and the native vaccine against this parasite, deeply affects the

rumen microbial communities, with severe consequences for their functionality. We have identified several potentially pathogenic genera with increased abundances in the rumen of infected animals, in comparison to uninfected, unvaccinated animals. Additionally, we have shown that microbial genes involved in the fermentation process reported in the literature as influencing the host animal's appetite, feed conversion efficiency, and methane production were altered in infected and vaccinated animals in comparison to uninfected ones. Rumen microbial genes enriched in infected animals were mainly associated with bacterial defence mechanisms and peptidoglycan biosynthesis, whereas those depleted were mainly involved in functions of environmental sensing and communication, which highlights a severe level of dysbiosis. Therefore, potential negative consequences of parasitism by nematodes in cattle, e.g., animal health and welfare issues, inappetence, and lower productivity, are not exclusively due to the direct effect of the nematode on the ruminant, i.e., by direct damage to gastric cells and alterations to abomasal pH, but also due to dysbiosis of the rumen microbiome. Additionally, we showed that the vaccinating the animals may have a positive effect on their rumen microbiome, by potentially preventing alterations of the microbiome associated with increased methane emissions, and hindered nitrogen utilization. However, the comparison of vaccinated with unvaccinated uninfected animals showed increased abundances of opportunistic pathogens. The differential results obtained from the pairwise comparisons of microbiome profiles of infected animals with high and low cumulative faecal egg count to those of uninfected animals indicated that microbiome features could be used as biomarkers for resilience of the host animal to infection by *O. ostertagi*, which could potentially be applied in the development of targeted dietary interventions (through pre- or probiotics) to alleviate the impact of the infection. Furthermore, we underlined the need to further develop anthelminthic vaccines that build on the hosts' immune response, without negatively impacting the rumen microbiome and associated functionality.

5.7 References

Allison, M. J., Bryant, M. P., and Doetsch, R. N. (1958). Volatile fatty acid growth factor for cellulolytic cocci of bovine rumen. *Science (80-.).* 128, 474–475. doi:10.1126/science.os-2.57.341-a.

Almería, S., Adelantado, C., Charlier, J., Claerebout, E., and Bach, A. (2009). Ostertagia ostertagi antibodies in milk samples: Relationships with herd management and milk production parameters in two Mediterranean production systems of Spain. *Res. Vet. Sci.* 87, 416–420. doi:10.1016/j.rvsc.2009.05.001.

Ames, B. N., Martin, R. G., and Garry, B. J. (1961). The first step of histidine biosynthesis. *J. Biol. Chem.* 236, 2019–2026. doi:10.1016/s0021-9258(18)64123-7.

Auffret, M. D., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., John Wallace, R., Freeman, T. C., et al. (2017). The rumen microbiome as a reservoir of antimicrobial resistance and pathogenicity genes is directly affected by diet in beef cattle. *Microbiome* 5, 159–169. doi:10.1186/s40168-017-0378-z.

Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C. A., Watson, M., and Roehe, R. (2020). Identification of microbial genetic capacities and potential mechanisms within the rumen microbiome explaining differences in beef cattle feed efficiency. *Front. Microbiol.* 11. doi:10.3389/fmicb.2020.01229.

Auffret, M. D., Stewart, R., Dewhurst, R. J., Duthie, C.-A., Rooke, J. A., Wallace, R. J., et al. (2018). Identification, comparison, and validation of robust rumen microbial biomarkers for methane emissions using diverse Bos taurus breeds and basal diets. *Front. Microbiol.* 8, 1–15. doi:10.3389/fmicb.2017.02642.

Badali, H., Chander, J., Gupta, A., Rani, H., Punia, R. S., De Hoog, G. S., et al. (2011). Fatal cerebral phaeohyphomycosis in an immunocompetent individual due to Thielavia subthermophila. *J. Clin. Microbiol.* 49, 2336–2341. doi:10.1128/JCM.02648-10.

Barron, G. L. (1987). “Fungal parasites and predators of rotifers, nematodes, and other invertebrates,” in *Biodiversity of fungi: inventory and monitoring methods* (Elsevier Inc.), 435–450. doi:10.1016/B978-0-12-509551-8.50022-2.

Berghen, P., Hilderson, H., Vercruysse, J., and Dorny, P. (1993). Evaluation of pepsinogen, gastrin and antibody response in diagnosing ostertagiasis. *Vet. Parasitol.* 46, 175–195. doi:10.1016/0304-4017(93)90057-T.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–63.

Campos, F. C., Filho, A. L. A., Corrêa, P. S., Nazato, C., Monnerat, R. G., Mcmanus, C. M., et al. (2019). Rumen degradability and gas production as influenced by different strains of Bacillus thuringiensis. *Can. J. Anim. Sci.* 954, 951–954.

Cani, P. D., and Knauf, C. (2016). How gut microbes talk to organs: The role of endocrine and nervous routes. *Mol. Metab.* 5, 743–752. doi:10.1016/j.molmet.2016.05.011.

Cano, J., Guarro, J., and Gené, J. (2004). Molecular and morphological identification of Colletotrichum species of clinical interest. *J. Clin. Microbiol.* 42, 2450–2454. doi:10.1128/JCM.42.6.2450-2454.2004.

Castillo-González, A. R., Burrola-Barraza, M. E., Domínguez-Viveros, J., and Chávez-Martínez, A. (2014). Rumen microorganisms and fermentation. *Arch. Med. Vet.* 46, 349–361. doi:10.4067/S0301-732X2014000300003.

Chacher, B., Wang, D. M., Liu, H. Y., and Liu, J. X. (2012). Degradation of L-arginine and N-carbamoyl glutamate and their effect on rumen fermentation in vitro. *Ital. J. Anim. Sci.* 11, 374–377. doi:10.4081/ijas.2012.e68.

Charlier, J., Dorny, P., Levecke, B., Demeler, J., von Samson-Himmelstjerna, G., Höglund, J., et al. (2011). Serum pepsinogen levels to monitor gastrointestinal nematode infections in cattle revisited. *Res. Vet. Sci.* 90, 451–456. doi:10.1016/j.rvsc.2010.06.029.

Courvalin, P. (2006). Vancomycin resistance in Gram-positive cocci. *Oxford Journals* 42.

da Silva, L. L., Moreno, H. L. A., Correia, H. L. N., Santana, M. F., and de Queiroz, M. V. (2020). Colletotrichum: species complexes, lifestyle, and peculiarities of some sources of genetic variability. *Appl. Microbiol. Biotechnol.* 104, 1891–1904. doi:10.1007/s00253-020-10363-y.

de Jong, S. E., Olin, A., and Pulendran, B. (2020). The impact of the microbiome on immunity to vaccination in humans. *Cell Host Microbe* 28, 169–179. doi:10.1016/j.chom.2020.06.014.

Dehority, B. A., Johnson, R. R., Bentley, O. G., and Moxon, A. L. (1958). Studies on the metabolism of valine, proline, leucine and isoleucine by rumen microorganisms in vitro. *Arch. Biochem. Biophys.* 78, 385–396. doi:10.1016/s0021-9258(18)71126-5.

Derosa, A. A., Chirgwin, S. R., Fletcher, J., Williams, J. C., and Klei, T. R. (2005). Exsheathment of Ostertagia ostertagi infective larvae following exposure to bovine rumen contents derived from low and high roughage diets. *Vet. Parasitol.* 129, 77–81. doi:10.1016/j.vetpar.2004.12.019.

Deutscher, J., Francke, C., and Postma, P. W. (2006). How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* 70, 939–1031. doi:10.1128/mmbr.00024-06.

Díaz-Sánchez, S., Perrotta, A. R., Rockafellow, I., Alm, E. J., Okimoto, R., Hawken, R., et al. (2019). Using fecal microbiota as biomarkers for predictions of performance in the selective breeding process of pedigree broiler breeders. *PLoS One* 14, e0216080. doi:10.1371/journal.pone.0216080.

Dimonaco, N. J., Salavati, M., and Shih, B. B. (2021). Computational analysis of sars-cov-2 and sars-like coronavirus diversity in human, bat and pangolin populations. *Viruses* 13, 1–22. doi:10.3390/v13010049.

Ding, L., Shen, Y., Wang, Y., Zhou, G., Zhang, X., Wang, M., et al. (2019). Jugular arginine supplementation increases lactation performance and nitrogen utilization efficiency in lactating dairy cows. *J. Anim. Sci. Biotechnol.* 10, 1–10. doi:10.1186/s40104-018-0311-8.

Donadio, S., and Sosio, M. (2009). “Glycopeptides, Antimicrobial,” in *Encyclopedia of Microbiology*, 455–471. doi:10.1016/B978-012373944-5.00040-7.

Farningham, D. A. H., Mercer, J. G., and Lawrence, C. B. (1993). Satiety signals in sheep: Involvement of CCK, propionate, and vagal CCK binding sites. *Physiol. Behav.* 54, 437–442. doi:10.1016/0031-9384(93)90232-5.

Flint, H. J., Duncan, S. H., Scott, K. P., and Louis, P. (2014). Links between diet, gut microbiota composition and gut metabolism. *Proc. Nutr. Soc.* 760, 13–22. doi:10.1017/S0029665114001463.

Fox, M. T. (1993). Pathophysiology of infection with Ostertagia ostertagi in cattle. *Vet. Parasitol.* 46, 143–158. doi:10.1016/0304-4017(93)90055-R.

Fox, M. T., Gerrelli, D., Pitt, S. R., Jacobs, D. E., Gill, M., and Gale, D. L. (1989a). Ostertagia ostertagi infection in the calf: effects of a trickle challenge on appetite, digestibility, rate of passage of digesta and liveweight gain. *Res. Vet. Sci.* 47, 294–298. doi:10.1016/s0034-5288(18)31249-9.

Fox, M. T., Gerrelli, D., Pitt, S. R., Jacobs, D. E., Hart, I. C., and Simmonds, A. D. (1987). Endocrine effects of a single infection with Ostertagia ostertagi in the calf. *Int. J. Parasitol.* 17, 1181–1185. doi:10.1016/0020-7519(87)90170-6.

Fox, M. T., Gerrelli, D., Shivalkar, P., and Jacobs, D. E. (1989b). Effect of omeprazole treatment on feed intake and blood gastrin and pepsinogen levels in the calf. *Res. Vet. Sci.* 46, 280–282. doi:10.1016/s0034-5288(18)31160-3.

Fox, M. T., Uche, U. E., Vaillant, C., Ganabadi, S., and Calam, J. (2002). Effects of Ostertagia ostertagi and omeprazole treatment on feed intake and gastrin-related responses in the calf. *Vet. Parasitol.* 105, 285–301. doi:10.1016/S0304-4017(02)00026-2.

Fox, N. J., Smith, L. A., Houdijk, J. G. M., Athanasiadou, S., and Hutchings, M. R. (2018). Ubiquitous parasites drive a 33% increase in methane yield from livestock. *Int. J. Parasitol.* 48, 1017–1021. doi:10.1016/j.ijpara.2018.06.001.

Francis, M. J., Doherty, R. R., Patel, M., Hamblin, J. F., Ojaimi, S., and Korman, T. M. (2011). Curtobacterium flaccumfaciens septic arthritis following puncture with a coxspur hawthorn thorn. *J. Clin. Microbiol.* 49, 2759–2760. doi:10.1128/JCM.00340-11.

Gómez-Garcés, J. L., Oteo, J., García, G., Aracil, B., Alós, J. I., and Funke, G. (2001). Bacteremia by Dermabacter hominis, a rare pathogen. *J. Clin. Microbiol.* 39, 2356–2357. doi:10.1128/JCM.39.6.2356-2357.2001.

Grant-Downton, R. T., Terhem, R. B., Kapralov, M. V., Mehdi, S., Rodriguez-Enriquez, M. J., Gurr, S. J., et al. (2014). A novel Botrytis species is associated with a newly emergent foliar disease in cultivated Hemerocallis. *PLoS One* 9. doi:10.1371/journal.pone.0089272.

Greenacre, M. (2018). “Chapter 3: Logratio transformations,” in *Compositional Data Analysis in Practice*, ed. C. & H. / C. Press., 17–24.

Greenwood, R. H., and Titgemeyer, E. C. (2000). Limiting amino acids for growing Holstein steers limit-fed soybean hull-based diets. *J. Anim. Sci.* 78, 1997–2004. doi:10.2527/2000.7871997x.

Haque, M. F. U., Xu, H. J., Colin Murrell, J., and Crombie, A. (2020). Facultative methanotrophs – diversity, genetics, molecular ecology and biotechnological potential: A mini-review. *Microbiol. (United Kingdom)* 166, 894–908. doi:10.1099/mic.0.000977.

Hawkins, J. A. (1993). Economic benefits of parasite control in cattle. *Vet. Parasitol.* 46, 159–173. doi:10.1016/0304-4017(93)90056-S.

Heizer, E., Zarlenga, D. S., Rosa, B., Gao, X., Gasser, R. B., De Graef, J., et al. (2013). Transcriptome analyses reveal protein and domain families that delineate stage-related development in the economically important parasitic nematodes, Ostertagia ostertagi and Cooperia oncophora. *BMC Genomics* 14. doi:10.1186/1471-2164-14-118.

Houdijk, J. G. M., Tolkamp, B. J., Rooke, J. A., and Hutchings, M. R. (2017). Animal health and greenhouse gas intensity: the paradox of periparturient parasitism. *Int. J. Parasitol.* 47, 633–641. doi:10.1016/j.ijpara.2017.03.006.

Houlden, A., Hayes, K. S., Bancroft, A. J., Worthington, J. J., Wang, P., Grencis, R. K., et al. (2015). Chronic Trichuris muris infection in C57BL/6 mice causes significant changes in host microbiota and metabolome: Effects reversed by pathogen clearance. *PLoS One* 10. doi:10.1371/journal.pone.0125945.

Islam, M., Kim, S. H., Son, A. R., Ramos, S. C., Jeong, C. D., Yu, Z., et al. (2021). Seasonal influence on rumen microbiota, rumen fermentation and enteric methane emissions of holstein and jersey steers under the same total mixed ration. *Animals* 11. doi:10.3390/ani11041184.

Jackson, C. R., Lombard, J. E., Dargatz, D. A., and Fedorka-Cray, P. J. (2011). Prevalence, species distribution and antimicrobial resistance of enterococci isolated from US dairy cattle. *Lett. Appl. Microbiol.* 52, 41–48. doi:10.1111/j.1472-765X.2010.02964.x.

Kandi, V., Palange, P., Vaish, R., Bhatti, A. B., Kale, V., Kandi, M. R., et al. (2016). Emerging bacterial infection: identification and clinical significance of Kocuria species. *Cureus* 8. doi:10.7759/cureus.731.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1016/j.meegid.2016.07.022.

Kenyon, F., and Jackson, F. (2012). Targeted flock/herd and individual ruminant treatment approaches. *Vet. Parasitol.* 186, 10–17. doi:10.1016/j.vetpar.2011.11.041.

Kim, J. S., Kim, W. S., Lee, K., Won, C. J., Kim, J. M., Eum, S. Y., et al. (2013). Differential immune responses to Segniliparus rotundus and Segniliparus rugosus infection and analysis of their comparative virulence profiles. *PLoS One* 8. doi:10.1371/journal.pone.0059646.

Kirby, R., Sangal, V., Tucker, N. P., Zakrzewska-Czerwińska, J., Wierzbicka, K., Herron, P. R., et al. (2012). Draft genome sequence of the human pathogen Streptomyces somaliensis, a significant cause of actinomycetoma. *J. Bacteriol.* 194, 3544–3545. doi:10.1128/JB.00534-12.

Lass, S., Hudson, P. J., Thakar, J., Saric, J., Harvill, E., Albert, R., et al. (2013). Generating super-shedders: Co-infection increases bacterial load and egg production of a gastrointestinal helminth. *J. R. Soc. Interface* 10. doi:10.1098/rsif.2012.0588.

Lebreton, F., Willems, R. J. L., and Gilmore, M. S. (2014). “Enterococcus diversity, origins in nature, and gut colonization,” in *Enterococci: From commensals to leading causes of drug resistant infection*, 5–64.

Li, F., and Guan, L. L. (2017). Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Appl. Environ. Microbiol.* 83. doi:10.1128/aem.00061-17.

Li, R. W., Li, W., Sun, J., Yu, P., Baldwin, R. L., and Urban, J. F. (2016). The effect of helminth infection on the microbial composition and structure of the caprine abomasal microbiome. *Sci. Rep.* 6, 1–10. doi:10.1038/srep20606.

Li, R. W., Wu, S., Li, W., Navarro, K., Couch, R. D., Hill, D., et al. (2012). Alterations in the porcine colon microbiota induced by the gastrointestinal nematode Trichuris suis. *Infect. Immun.* 80, 2150–2157. doi:10.1128/IAI.00141-12.

Li, Y. Q., Xi, Y. M., Wang, Z. D., Zeng, H. F., and Han, Z. (2020). Combined signature of rumen microbiome and metabolome in dairy cows with different feed intake levels. *J. Anim. Sci.* 98, 1–15. doi:10.1093/JAS/SKAA070.

Lima, J., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C.-A., Snelling, T. J., et al. (2019). Identification of rumen microbial genes involved in pathways linked to appetite, growth, and feed conversion efficiency in cattle. *Front. Genet.* 10:701. doi:10.3389/fgene.2019.00701.

Martínez-Álvaro, M., Auffret, M. D., Duthie, C.-A., Dewhurst, R. J., Cleveland, M., Watson, M., et al. (2021). Bovine host genome acts on specific metabolism, communication and genetic processes of rumen microbes host-genomically linked to methane emissions. *Res. Sq. - Prepr.*, 0–37.

Martínez-Álvaro, M., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., et al. (2020). Identification of complex rumen microbiome interaction within diverse functional niches as mechanisms affecting the variation of methane emissions in bovine. *Front. Microbiol.* 11:659. doi:10.3389/fmicb.2020.00659.

McCuistion, K. C., Titgemeyer, E. C., Awawdeh, M. S., and Gnad, D. P. (2004). Histidine utilization by growing steers is not negatively affected by increased supply of either ammonia or amino acids. *J. Anim. Sci.* 82, 759–769. doi:10.1093/ansci/82.3.759.

Menahan, L. A., and Schultz, L. H. (1964). Metabolism of leucine and valine within the rumen. *J. Dairy Sci.* 47, 1080–1085. doi:10.3168/jds.S0022-0302(64)88849-4.

Meyvis, Y., Geldhof, P., Gevaert, K., Timmerman, E., Vercruysse, J., and Claerebout, E. (2007). Vaccination against Ostertagia ostertagi with subfractions of the protective ES-thiol fraction. *Vet. Parasitol.* 149, 239–245. doi:10.1016/j.vetpar.2007.08.014.

Mitsumori, M., Ajisaka, N., Tajima, K., Kajikawa, H., and Kurihara, M. (2002). Detection of Proteobacteria from the rumen by PCR using methanotroph-specific primers. *Lett. Appl. Microbiol.* 35, 251–255. doi:10.1046/j.1472-765X.2002.01172.x.

Muñoz, R., García, J. L., Carrascosa, A. V., and Gonzalez, R. (2004). Cloning of the authentic bovine gene encoding pepsinogen A and its expression in microbial cells. *Appl. Environ. Microbiol.* 70, 2588–2595. doi:10.1128/AEM.70.5.2588-2595.2004.

Murray, M. (1969). Structural changes in bovine ostertagiasis associated with increased permeability of the bowel wall to macromolecules. *Gastroenterology* 56, 763–772. doi:10.1016/S0016-5085(69)80039-9.

Myers, G. H., and Taylor, R. F. (1989). Ostertagiasis in cattle. *J. Vet. Diagnostic Investig.* 1, 195–200. doi:10.1177/104063878900100225.

Nagpal, R., Shrivastava, B., Kumar, N., Dhewa, T., and Sahay, H. (2015). “Microbial Feed Additives,” in *Rumen Microbiology: From Evolution to Revolution* (Springer), 161–176. doi:DOI 10.1007/978-81-322-2401-3.

Nowland, T. L., Plush, K. J., Barton, M., and Kirkwood, R. N. (2019). Development and function of the intestinal microbiome and potential implications for pig production. *Animals* 9, 1–15. doi:10.3390/ani9030076.

Onodera, R. (2003). Essentiality of histidine in ruminant and other animals including human beings. *Asian-Australasian J. Anim. Sci.* 16, 445–454. doi:10.5713/ajas.2003.445.

Palevich, N., Kelly, W. J., Leahy, S. C., Denman, S., Altermann, E., Rakonjac, J., et al. (2020). Comparative genomics of rumen Butyrivibrio spp. uncovers a continuum of polysaccharide-degrading capabilities. *Appl. Environ. Microbiol.* 86, 1–19. doi:10.1128/AEM.01993-19.

Parkins, J. J., Bairden, K., and Armour, J. (1982). Ostertagia ostertagi in calves: a growth, nitrogen balance and digestibility study conducted during winter feeding following thiabendazole anthelmintic therapy. *J. Comp. Path* 92, 219–227.

Plieskatt, J. L., Deenonpoe, R., Mulvenna, J. P., Krause, L., Sripa, B., Bethony, J. M., et al. (2013). Infection with the carcinogenic liver fluke Opisthorchis viverrini modifies intestinal and biliary microbiome. *FASEB J.* 27, 4572–4584. doi:10.1096/fj.13-232751.

Prins, R. A., van Vugt, F., Hungate, R. E., and van Vorstenbosch, C. J. A. H. V. (1972). A comparison of strains of Eubacterium cellulosolvens from the rumen. *Antonie Van Leeuwenhoek* 38, 153–161. doi:10.1007/BF02328087.

Rebets, Y., Tokovenko, B., Lushchyk, I., Rückert, C., Zaburannyi, N., Bechthold, A., et al. (2014). Complete genome sequence of producer of the glycopeptide antibiotic Aculeximycin Kutzneria albida DSM 43870T, a representative of minor genus of Pseudonocardiaceae. *BMC Genomics* 15, 1–15. doi:10.1186/1471-2164-15-885.

Roehe, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., McKain, N., Ross, D. W., et al. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* 12, e1005846. doi:10.1371/journal.pgen.1005846.

Saverwyns, H. (2008). Study of Ostertagia ostertagi excretory-secretory products. Available at: https://core.ac.uk/download/pdf/55748002.pdf.

Schmidt, H., and Hensel, M. (2004). Pathogenicity Islands in Bacterial Pathogenesis. *Clincal Microbiol. Rev.* 17, 14–56. doi:10.1128/CMR.17.1.14.

Segers, R., Butt, T. M., Kerry, B. R., and Peberdy, J. F. (1994). The nematophagous fungus Verticillium chlamydosporium produces a chymoelastase-like protease which hydrolyses host nematode proteins in situ. *Microbiology* 140, 2715–2723. doi:10.1099/00221287-140-10-2715.

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

Shabat, S., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., Berg Miller, M. E., et al. (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J.* 10, 2958–2972. doi:10.1038/ismej.2016.62.

Shapiro, C. L., Haft, R. F., Gantz, N. M., Doern, G. V., Christenson, J. C., Richard, O., et al. (1992). Tsukamurella paurometabolum: a novel pathogen causing catheter-related bacteremia in patients with cancer. *Clin. Infect. Dis.* 14, 200–203. doi:10.1093/clinids/14.1.200.

Skipper, C., Ferrieri, P., and Cavert, W. (2020). Bacteremia and central line infection caused by Bosea thiooxidans. *IDCases* 19, e00676. doi:10.1016/j.idcr.2019.e00676.

Stergiadis, S., Cabeza-Luna, I., Mora-Ortiz, M., Stewart, R. D., Dewhurst, R. J., Humphries, D. J., et al. (2021). Unravelling the role of rumen microbial communities, genes, and activities on milk fatty acid profile using a combination of omics approaches. *Front. Microbiol.* 11, 1–15. doi:10.3389/fmicb.2020.590441.

Stewart, R. D., Auffret, M. D., Warr, A., Wiser, A. H., Press, M. O., Langford, K. W., et al. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* 9, 870. doi:10.1038/s41467-018-03317-6.

Strobel, T., Al-Dilaimi, A., Blom, J., Gessner, A., Kalinowski, J., Luzhetska, M., et al. (2012). Complete genome sequence of Saccharothrix espanaensis DSM 44229 T and comparison to the other completely sequenced Pseudonocardiaceae. *BMC Genomics* 13. doi:10.1186/1471-2164-13-465.

Suzuki, H., Arshava, E. V., Ford, B., and Nauseef, W. M. (2019). Don't let its name fool you: Relapsing thoracic actinomycosis caused by pseudopropionibacterium propionicum (formerly propionibacterium propionicum). *Am. J. Case Rep.* 20, 1961–1965. doi:10.12659/AJCR.919775.

Taguchi, H., Koike, S., Kobayashi, Y., Cann, I. K. O., and Karita, S. (2004). Partial characterization of structure and function of a xylanase gene from the rumen hemicellulolytic bacterium Eubacterium ruminantium. *Anim. Sci. J.* 75, 325–332. doi:10.1111/j.1740-0929.2004.00193.x.

Theoulakis, P., Goldblum, D., Zimmerli, S., Muehlethaler, K., and Frueh, B. E. (2009). Keratitis resulting from Thielavia subthermophila Mouchacca. *Cornea* 28, 1067–1069. doi:10.1097/ICO.0b013e31819717f4.

Ventura, A. K., Beauchamp, G. K., and Mennella, J. A. (2012). Infant regulation of intake: The effect of free glutamate content in infant formulas. *Am. J. Clin. Nutr.* 95, 875–881. doi:10.3945/ajcn.111.024919.

Wang, L., Liu, K., Wang, Z., Bai, X., Peng, Q., and Jin, L. (2019). Bacterial community diversity associated with different utilization efficiencies of nitrogen in the gastrointestinal tract of goats. *Front. Microbiol.* 10, 1–14. doi:10.3389/fmicb.2019.00239.

Wellmann, R., Bennewitz, J., Preuss, S., Camarinha-Silva, A., Vital, M., and Maushammer, M. (2017). Host genome influence on gut microbial composition and microbial prediction of complex traits in pigs. *Genetics* 206, 1637–1644. doi:10.1534/genetics.117.200782.

Winkler, M. E., and Ramos-Montanez, S. (2009). Biosynthesis of Histidine. *EcoSal Plus.* 3, 1–18. doi:10.1128/ecosalplus.3.6.1.9.Biosynthesis.

Wirth, R., Kádár, G., Kakuk, B., Maróti, G., Bagi, Z., Szilágyi, Á., et al. (2018). The planktonic core microbiome and core functions in the cattle rumen by next generation sequencing. *Front. Microbiol.* 9, 1–19. doi:10.3389/fmicb.2018.02285.

Won, M. Y., Oyama, L. B., Courtney, S. J., Creevey, C. J., and Huws, S. A. (2020). Can rumen bacteria communicate to each other? *Microbiome* 8, 1–8. doi:10.1186/s40168-020-00796-y.

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

Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36, 808–812. doi:10.2144/04365st04.

Chapter 6 The caecal microbiome profiles are affected by the presence of the abomasal parasitic nematode *Ostertagia ostertagi* in dairy cattle

6.1 Abstract

Ostertagia ostertagi is one of the most economically impactful gastrointestinal parasitic nematodes in cattle production systems, causing inappetence, reduced growth rates, and even death. Since the microbiome in bovine gastrointestinal tract (GIT) is so closely associated with host production and health traits, we investigated the influence of the nematode on the caecal microbiome based on animals challenged by oral administration of 1000 L3 infectious *O. ostertagi* larvae/day for 25 days after receiving either a native vaccine against the parasite or an injection of adjuvant-only (positive control). The caecal microbiome samples of 16 cattle collected at slaughter were whole metagenome sequenced and resolved into microbial taxa and microbial genes compositions. The caecal microbiome at the taxonomic and genetic levels of infected and vaccinated animals was compared to that of uninfected animals (negative control), using an iterative partial least squares-based analyses. The results showed that vaccinated animals were depleted of e.g., opportunistic pathogen *Bordetella*. Additionally, the infection and the vaccination affected microbial genes mostly associated with bacterial cell wall and environmental sensing, including ABC transporters and genes involved in the two-systems components, and O-antigen nucleotide sugar biosynthesis. The caecum of infected animals was enriched in microbial genes associated with degradation of aromatic compounds. Furthermore, comparing the microbiome of infected with that of vaccinated animals revealed that the vaccine leads to depletion of microbial genes e.g., *cheA* (two-component system, chemotaxis family), previously associated with improved feed conversion efficiency in beef cattle, and microbial genes associated with production of bacteriocins, e.g., lantibiotic biosynthesis proteins *nisB* and *nisC*. This study revealed that the abomasal nematode *O. ostertagi* substantially affected the caecal microbiome and

identified potential microbial biomarkers that could be used for selection of animals in breeding programmes, or the development of individualized nutritional interventions to circumvent the impacts of caecal microbiome dysbiosis on the hosts' performance.

6.2 Introduction

The symbiotic relationship between the gastrointestinal microbiome and the bovine host is a main focus of research, due to its involvement in the animal's health (O'Hara et al., 2020), productivity (Guan et al., 2008; Lima et al., 2019; O'Hara et al., 2020), and environmental impact (Difford et al., 2018; Martínez-Álvaro et al., 2020). Most often, the rumen and its microbiota are the subject of the interest due to their essential role in the fermentation of plant biomass. The fermentation of complex polysaccharides such as cellulose and hemicellulose, into volatile fatty acids, microbial protein, and vitamins by the resident flora has been shown to contribute 62% of the total metabolizable energy (ME) in steers (Siciliano-Jones and Murphy, 1989). However, the hindgut also plays a part in reutilizing previously undigested nutrients and the caecum fermentation has been reported to contribute 9% of the ME (Siciliano-Jones and Murphy, 1989). The rumen and the caecum differ in many aspects, e.g., their location within the gastrointestinal tract (GIT), with the rumen being the first fermentation chamber in the gastrointestinal tract of ruminants, and the caecum being the first segment of the large intestine, which influence the residing flora. Whereas the rumen is dominated by Bacteroidetes (cellulolytic and saccharolytic bacteria (Naas et al., 2014), characteristically associated with high-energy diets (Ottman et al., 2012)), higher abundances of Firmicutes (mostly cellulolytic) have been reported in the caecum microbiota (de Oliveira et al., 2013). The specific microbiome of each GIT organ has been suggested to be associated with different functions; for example, microbial protein produced in the rumen flows into the abomasum, where it is digested and absorbed to be used as main protein source for the ruminant (Strom and Øskov, 1984). On the other hand, the caecum microbiota is involved in the fermentation of available nutrients that bypass the rumen and the abomasum

(de Oliveira et al., 2013). With 70% of the operational taxonomic units (OTUs) identified in the caecum being associated with increased growth rates in cattle, the caecal microbiome has been suggested as having an underappreciated influence on the host's performance (Freetly et al., 2020).

The abomasal nematode *Ostertagia ostertagi* is a cattle parasite that impairs the gastrointestinal function of the host (usually first and/or second season calves), by negatively influencing the host's appetite and liveweight gain (Fox et al., 1989), decreasing rate of passage of digesta (Fox et al., 1989), and by altering gut motility (Bueno et al., 1982; Fox et al., 2006). During the parasitic life stage of *O. ostertagi*, the third stage infective larvae (L3) invade the gastric glands in the abomasum, where they develop into fourth stage larvae (L4) and into young adults (L5) that exit the gastric glands, damaging them. The damaged gastric glands have been associated with reduced hydrochloric acid production (due to damaged parietal cells), and subsequently to increased pH observed in the abomasum of infected animals, accompanied by hypergastrinemia (Fox et al., 1987), and leading to an accumulation of non-activated pepsinogen in the gastric glands and increased permeability of the mucosa (allowing the pepsinogen to leak back into the blood stream) (Fox et al., 2002). Increased abomasal pH, blood gastrin, and pepsinogen levels are also associated with the presence of the nematode (by transplantation of adult worms, i.e., with no damage to the gastric glands, McKellar et al., 1986). Alterations of the abomasal pH have been shown to critically affect the abomasal microbiome. For example, the increased pH in the abomasum could increase the survivability of Gram-negative pathogens ingested with the feed, such as *Salmonella* and *Escherichia coli* (Constable et al., 2006).

In comparison to other bovine pathogens, immunity to *O. ostertagi* takes a long time to develop, with some calves remaining susceptible for up to two years of age, and is incomplete, with *Ostertagia* being reported as suppressing host cellular and antibody responses to infection (Klesius, 1988). The prevention and control of the parasite in bovine herds is most often achieved by the use

of anthelmintics, but the drawback of resistance development (Geerts et al., 1987; Edmonds et al., 2010) makes vaccination a more attractive strategy. Considering the close relationship between the GIT microbiome and the host animals' performance traits, the ideal vaccine should build up the animal's immunity without negatively impacting the microbiome.

The main objective of the present study was to investigate the impact of abomasal *O. ostertagi* infection on the caecal microbiome of sub clinically infected animals. In addition, we explored the influence of a native vaccine on the caecal microbiome of dairy cattle. Furthermore, we elucidate the differences between the caecal and ruminal microbiome profiles and their functionality on healthy (i.e., uninfected) animals, using microbiome community and functional microbial gene profiles.

6.3 Materials and methods

6.3.1 Ethics statement

Immunizations and parasite challenges of cattle were performed at Moredun Research Institute (MRI) under Home Office licence 70/7914. Ethical approval was obtained from the MRI Animal Welfare and Ethical Review Body (E12/18). Animals were euthanized at the MRI post-mortem facility.

6.3.2 Animals and experimental procedure

An experimental trial was carried out in 2018 to determine the effect of a native vaccine against *Ostertagia ostertagi* in dairy cattle (British Friesian and Norwegian Red).

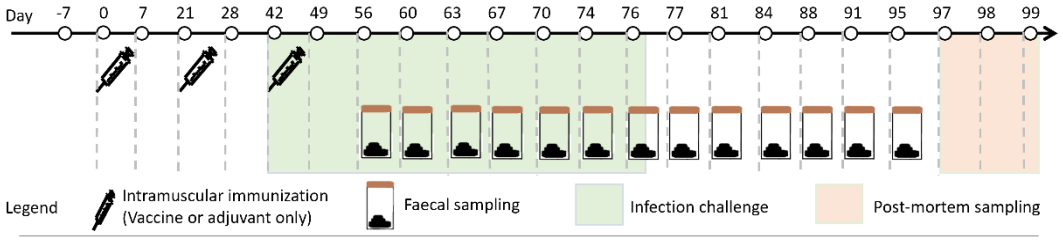


Figure 1. *Schematic representation of the experimental timeline.*

This experiment included 20 male calves of 4-5 months of age. Animals were allocated into 2 groups, balanced for breed and weight. Animals in group 1 were vaccinated using the native *O. ostertagi* vaccine, whereas animals in group 2 were treated with adjuvant-only (Quil A® (Brenntag Biosector) - 750ug per dose, positive control). Administration of vaccine or adjuvant-only occurred on days 0, 21 and 42 of the experimental trial (Figure 1). The infection challenge started on day 42, when animals were orally administered 1000 infectious L3 larvae per day, for 25 days. During the experimental trial, 13 faeces samples were collected per animal, and animals were evaluated for their cumulative faecal egg count (cFEC). The cFEC is an indicator of worm fitness, and it has been previously used to determine vaccine efficacy against *O. ostertagi* (Meyvis et al., 2007). A total of 12 animals were identified for post-mortem rumen and caecum digesta sampling: 8 animals from group 2 (infected non-vaccinated) were identified based on their high (n=4) and low (n=4) cFEC (CHE and CLE, respectively) and 4 animals from group 1 (infected and vaccinated, VAC) were selected from the cFEC-boxplot's second and third quantiles. Additionally, 4 calves used as negative control (uninfected and non-vaccinated, UNF) throughout the experimental trial were also included for rumen and caecum digesta sampling.

6.3.3 Sampling of ruminal and caecal digesta and whole metagenomic sequencing

Animals included in the experimental trial were slaughtered at the Moredun Research Centre. Sixteen samples of rumen and caecum digesta each were

collected within 10 minutes of the slaughter. DNA was extracted from the samples following the methodology described in (Yu and Morrison, 2004).

Illumina TruSeq DNA Nano libraries were prepared from genomic DNA and sequenced on Illumina NovaSeq 6000 systems by Edinburgh Genomics (Edinburgh, UK). Paired-end reads (2 × 150 bp) were generated, resulting in between 8 and 50 GB per sample (between 28 and 165 million paired reads). For taxonomic classification, the sequence reads of the samples were aligned to a database including genomes from the Hungate 1000 Collection (Seshadri et al., 2018) and metagenome-assembled genomes (MAGs) from beef rumen samples (Stewart et al., 2018) using Kraken (Wood and Salzberg, 2014). In total, 1200 genera found in all animals were identified and described as the genus having the highest similarity with the identified microbial genome or MAG. For functional annotation, DIAMOND was used to blast the reads for each sample against the KEGG database (downloaded 15/09/18) (Buchfink et al., 2015). Gene abundance was calculated as the sum of reads mapping to each KEGG orthologue.

6.3.4 Statistical analyses

6.3.4.1 Comparison of caecal and ruminal microbiome profiles

To assess the differences between the caecal and ruminal microbiome profiles at the domain level, we estimated the effect of the sample type (i.e., caecum and rumen) on the Archaea:Bacteria ratio in an ANOVA. Additionally, at the phylum and genera levels, we regressed the relative abundances of each of the five most abundant phyla and of the ten most abundant genera on sample type. The relative abundances of these taxa were used to create two stacked bar charts, in R.

Alpha and beta diversity at phylum, family, genus, and microbial genes levels were estimated within the caecum and the rumen microbiomes and compared using ANOVA, to further evaluate differences between the microbiome of these two organs. Alpha diversity was assessed through observed richness (Sobs), Shannon index (H’), and adjusted Shannon index (H’adj), whereas

beta diversity was assessed using Bray Curtis dissimilarity (BC). These indices were estimated using the vegan package (Oksanen et al., 2019) in R.

A total of 45 phyla, 461 families, 1200 genera, and 8723 microbial genes (i.e., KEGG orthologues) were identified in the 32 samples (16 samples from each rumen and caecum), and the relative abundances of each microbiome feature was calculated within sample. Microbial taxa and microbial genes that were absent from at least one sample or that had average relative abundance lower than 0.001% were removed from the datasets, leaving a total of 36 phyla, 367 families, 1059 genera and 3169 microbial genes for further analyses. The original datasets (i.e., read counts) were centred-log ratio (CLR) transformed, due to the compositional nature intrinsic to microbiome datasets (Greenacre, 2018). Partial least squares discriminant analyses (PLS-DA) were applied to CLR-transformed datasets at each taxonomic (and microbial genes) level to evaluate whether the microbiome profiles of UNF animals were different between sample types (i.e., rumen and caecum). These analyses were carried out using 'mixOmics' package (Cao et al., 2020) in R.

6.3.4.2 Comparing caecal microbiome profiles of unvaccinated challenged and vaccinated challenged with those of uninfected animals

We compared the alpha diversity within the caecal microbiome (at taxonomic and genetic level) of animals subjected to the different treatments, based on estimates of Sobs, H' and H'adj. Beta diversity was assessed and compared based on BC within treatment group. These indices were calculated using the vegan package (Oksanen et al., 2019) and compared using ANOVA in R.

At the genus-level, the original dataset included the sequence read counts of 1200 microbial genera. Microbial genera that were absent from at least 30% of the animals were removed from the dataset. The remaining zeros were imputed using the cmultRepl function of the package zCompositions (Palarea-Albaladejo and Martín-Fernández, 2020) in R (Version 1.4.1103). A total of 1194 microbial genera were used for further analyses, after CLR-transforming the datasets. At the microbial genes level, the original dataset included the

sequence read counts of 8484 microbial gene orthologs (i.e., KEGG) and the same thresholds and transformations were applied, resulting in a dataset containing the CLR-transformed abundance of 6647 microbial genes for further analyses.

To unravel the effect of the infection and the vaccine on the caecal microbiome, pairwise discriminant partial least squares models (PLS-DA) were calculated based on the CLR-transformed abundance of 1194 microbial genera and 6315 microbial genes identified, comparing CHE, CLE, and VAC, to UNF. After calculating the first model, the variables with variable importance in projection (VIP) scores equal or superior to 1 were selected, and used in a second PLS-DA. This process continued iteratively until we identified the smallest set of variables that best separated the treatment groups. Model quality was assessed through prediction error rates and area under the curve of operator characteristics (AUROC). Microbial genera and microbial genes identified as important for the discrimination between groups were also used in principal component analyses (PCA) as to evaluate their discriminative ability when no grouping effect is included in the model. Significant discrimination was assumed when (i) PLS-DA-based AUROC was equal to 1 with corresponding p-value < 0.05, (ii) no overlap between the 95% confidence ellipses in the score plot and the prediction error rate (based on confusion matrix) was the lowest and (iii) PCA-based score plot showed no overlap between the 95% confidence ellipses. We also compared the microbiome profiles of CHE and CLE to those of VAC, to better understand the impact of using a vaccine against the nematode. All analyses were performed using R Version 1.4.1103 (R Core Team, 2021).

6.4 Results

6.4.1 Differences between ruminal and caecal microbiota profiles

At domain level, we compared the Archaea:Bacteria ratio, which did not significantly differ between caecum and rumen (0.039 ± 0.027, and 0.028 ± 0.006, respectively).

6.4.1.1 Diversity indices in the caecum and rumen microbiome profiles

Diversity indices of observed richness (Sobs), Shannon index (H'), adjusted Shannon index (H'adj), and Bray Curtis dissimilarity (BC), calculated within sample type) were compared at phylum, family, genus, and microbial genes levels (Table 1).

Table 1. *Diversity at phylum, family, genus, and microbial genes levels in caecal and ruminal samples*

Taxon/Genes	Diversity index	Caecum	Rumen	p-value
Phylum	Sobs	44.813 ± 0.403	44.75 ± 0.447	0.681
	H'	1.093 ± 0.061	1.134 ± 0.053	0.050
	H'adj	0.287 ± 0.016	0.298 ± 0.015	0.051
	BC	0.13 ± 0.039	0.125 ± 0.082	0.823
Family	Sobs	460.313 ± 0.704	460.563 ± 0.512	0.260
	H'	2.549 ± 0.159	2.055 ± 0.145	3.17E-10
	H'adj	0.416 ± 0.026	0.335 ± 0.024	3.08E-10
	BC	0.177 ± 0.039	0.165 ± 0.071	0.567
Genus	Sobs	1195.625 ± 1.544	1194.875 ± 2.63	0.333
	H'	3.714 ± 0.302	2.816 ± 0.16	1.42E-11
	H'adj	0.524 ± 0.043	0.397 ± 0.023	1.42E-11
	BC	0.214 ± 0.036	0.167 ± 0.06	0.012
Microbial genes	Sobs	7146.625 ± 316.675	6937.438 ± 298.075	0.064
	H'	7.228 ± 0.039	7.25 ± 0.011	0.041
	H'adj	0.815 ± 0.003	0.82 ± 0.004	4.11E-4
	BC	0.086 ± 0.043	0.115 ± 0.082	0.224

Sobs, H', H'adj, and BC refer to observed richness, Shannon index, adjusted Shannon index, and Bray Curtis dissimilarity (calculated within sample type, caecum or rumen), respectively. Averages ± standard deviations are presented for each sample type; the p-values were obtained using ANOVA. Shaded cells represent significant differences.

No significant differences were observed in richness and Bray-Curtis dissimilarity between rumen and caecum. For unadjusted and adjusted

Shannon indices, we observed significantly higher evenness of the microbial community at family and genus levels in the caecum, whereas at microbial genes level, evenness was significantly higher in the rumen.

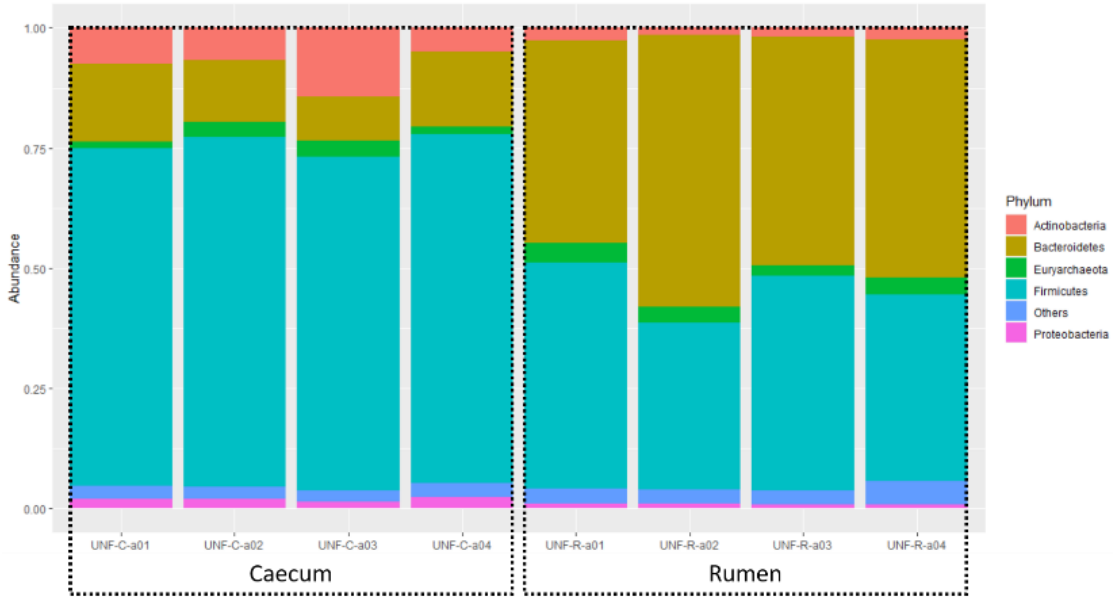


Figure 2. *Relative abundances of the 5 most abundant phyla in caecum and rumen. "Others" includes all phyla with relative abundance lower than 1%. UNF represents "uninfected" animals, C and R represent "caecum" and "rumen", respectively.*

The ANOVA analysis of the most abundant phyla in the rumen and caecum of UNF animals showed that Actinobacteria and Firmicutes were significantly more abundant in the caecum, whereas Bacteroidetes and Proteobacteria were significantly more abundant in the rumen (Figure 2, Table 2).

Table 2. *Average relative abundances of most abundant phyla in rumen and caecum.*

Phylum	Rumen	Caecum	p-value
Actinobacteria	2.1E-2 ± 5.2E-3	8.4E-2 ± 4.1E-2	2.2E-2
Bacteroidetes	4.9E-1 ± 5.9E-2	1.4E-1 ± 3.1E-2	4.3E-5
Euryarchaeota	3.2E-2 ± 7.9E-3	2.3E-2 ± 1.0E-2	2.2E-1
Firmicutes	4.1E-1 ± 5.6E-2	7.1E-1 ± 1.7E-2	5.0E-5
Proteobacteria	9.1E-3 ± 1.0E-3	2.0E-2 ± 4.0E-3	2.2E-3
Others	3.5E-2 ± 1.0E-2	2.7E-2 ± 2.7E-3	1.6E-1

The columns headings “Rumen” and “Caecum” contain the average ± the standard deviation of the relative abundance of each phylum in the samples obtained from uninfected animals. The p-value was obtained using ANOVA. Shaded cells represent significant differences (p-value<0.05).

The analyses of the most abundant genera in the microbiota profiles of uninfected animals in each sample type showed that *Bacteroides*, *Clostridium*, *Disctyostelium*, *Ruminococcus*, and *Sarcina* were significantly more abundant in the caecum, whereas *Butyrivibrio*, *Fibrobacter*, *Methanobrevibacter*, *Prevotella* and *Succiniclasticum* were significantly more abundant in the rumen (Figure 3, Table 3).

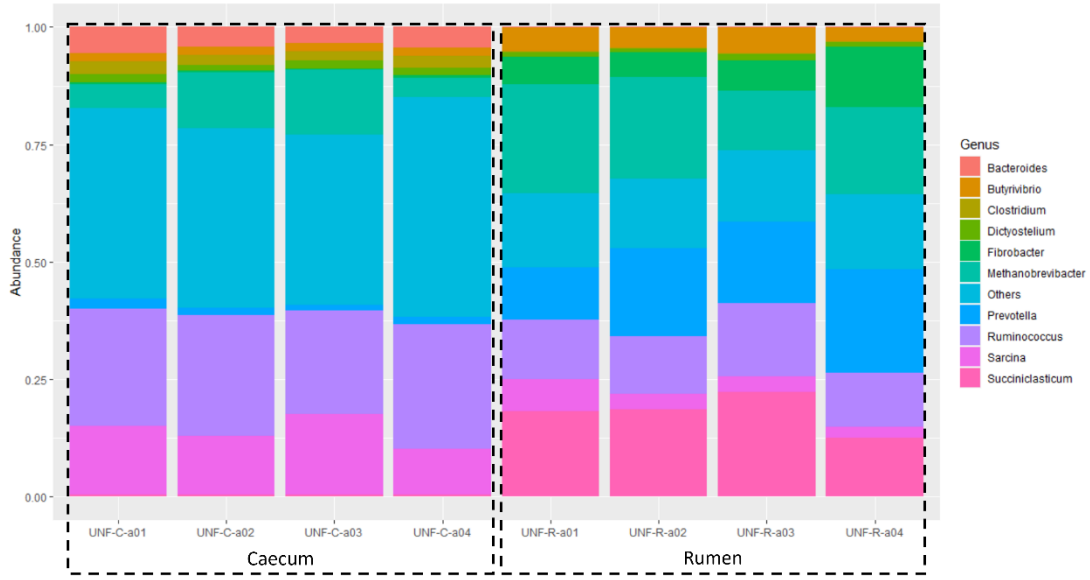


Figure 3. *Relative abundances of the 10 most abundant genera in caecum and rumen. "Others" includes all other genera. UNF represents "uninfected" animals, C and R represent "caecum" and "rumen", respectively.*

Table 3. *Average relative abundances of most abundant genera in rumen and caecum*

Genus	Rumen	Caecum	p-value	More abundant in
Bacteroides	1.1E-3 ± 1.6E-4	4.4E-2 ± 9.0E-3	8.1E-5	Caecum
Clostridium	1.7E-3 ± 1.3E-4	2.4E-2 ± 2.5E-3	2.3E-6	
Dictyostelium	1.0E-2 ± 3.3E-3	1.6E-2 ± 3.3E-3	3.8E-2	
Ruminococcus	1.3E-1 ± 1.8E-2	2.5E-1 ± 1.9E-2	1.0E-4	
Sarcina	1.4E-1 ± 1.9E-2	1.4E-1 ± 3.1E-2	1.8E-3	
Others	1.5E-1 ± 5.7E-3	4.1E-1 ± 4.6E-2	3.9E-5	
Butyrivibrio	4.5E-2 ± 1.2E-2	1.8E-2 ± 8.1E-4	3.6E-3	Rumen
Fibrobacter	7.5E-2 ± 3.6E-2	3.9E-3 ± 2.1E-3	7.2E-3	
Methanobrevibacter	1.9E-1 ± 4.7E-2	8.7E-2 ± 4.9E-2	2.2E-2	
Prevotella	1.7E-1 ± 4.6E-2	1.6E-2 ± 3.8E-3	4.9E-4	
Succiniclasticum	1.8E-1 ± 4.0E-2	3.7E-3 ± 1.2E-4	1.3E-4	

The columns named "Rumen" and "Caecum" contain the average ± the standard deviation of the relative abundance of each genus in the samples obtained from uninfected animals. The p-value was obtained from ANOVA comparison; all comparisons showed significant differences.

6.4.1.2 Comparison of ruminal and caecal microbiome profiles

PLS-DA analyses at phylum, family, genus, and microbial genes level (including 36, 367, 1059, and 3169 explanatory variables, respectively)

revealed significant discrimination between microbiome profiles obtained from rumen and caecum samples (Figure 4).

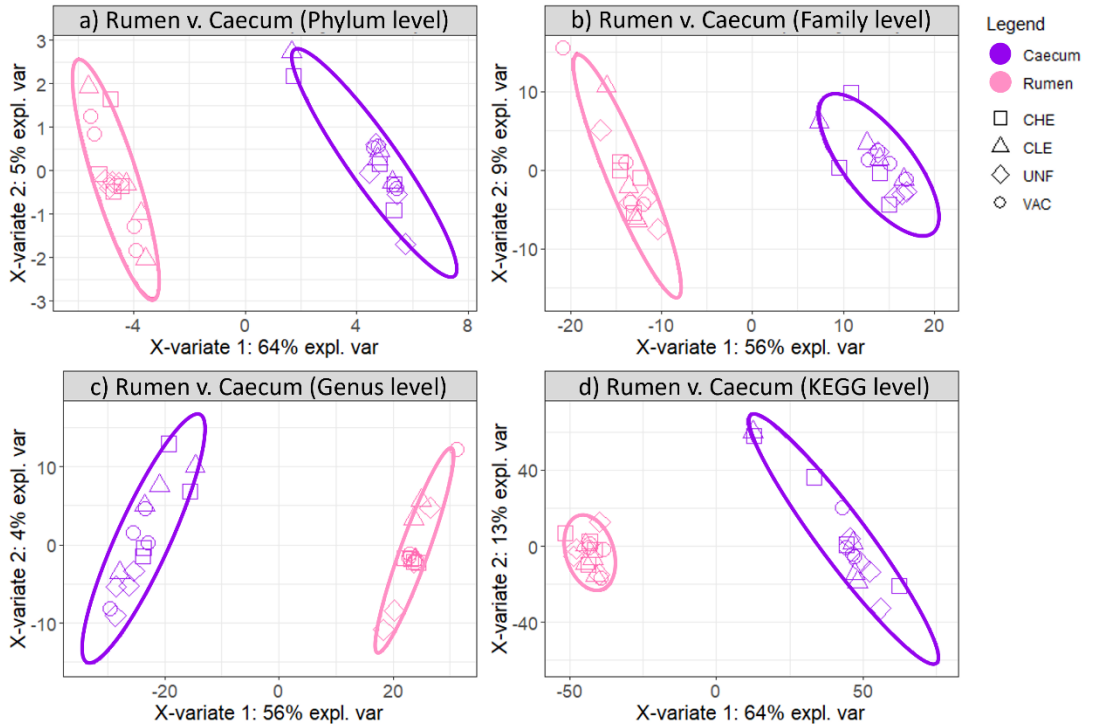


Figure 4. *Partial least squares discriminant analyses (PLS-DA) plots of all individuals based on a) 36 phyla, b) 367 families, c) 1059 genera, and d) 3169 microbial genes used in the discrimination analyses rumen and caecum. CHE, CLE, UNF and VAC refer to treatment groups infected showing high and low cumulative faecal egg count, uninfected, and vaccinated, respectively.*

PLS-DA revealed significant differences between the microbiome profiles obtained from rumen and caecum; whereas *Fibrobacter* and *Succiniclasticum* were more than 2,000-fold and *Lentisphaera* and *Prevotella* more than 1000-fold more abundant in the rumen, *Terrisporobacter*, *Methanocorpusculum*, *Clostridioides* and *Turicibacter* were more than 10,000-fold more abundant in the caecum.

At the functional level, this comparison revealed that 1816 microbial genes had VIP ≥ 1, 798 of which were enriched in the rumen, whereas 1018 were enriched in the caecum. For example, microbial genes *phoN*, and GALC (i.e.,

acid phosphatase (class A), and galactosylceramidase, respectively) were more than 3,000-fold more abundant in the rumen, whereas *agaW*, and *agaE* (i.e., N-acetylgalactosamine PTS system EIIC component, and N-acetylgalactosamine PTS system EIID component, respectively). Microbial genes enriched in the caecum were involved in carbon metabolism, biosynthesis of amino acids, and biosynthesis of cofactors pathways. At the module level, these microbial genes were found to participate in modules such as methanogenesis, acetate => methane (methane and carbon metabolism pathways); glycolysis (Embden-Meyerhof pathway), glucose => pyruvate (glycolysis/gluconeogenesis and carbon metabolism pathways); Lysine biosynthesis, acetyl-DAP pathway, aspartate => lysine (lysine biosynthesis and biosynthesis of amino acids pathways).

Microbial genes enriched in the rumen included ABC transporters and microbial genes involved in amino sugar and nucleotide sugar metabolism, biosynthesis of amino acids, and the two-component system pathways. At module level, the analyses of these microbial genes revealed that they had functions in the KDO2-lipid A biosynthesis, Raetz pathway, non-LpxL-LpxM type, and the KDO2-lipid A biosynthesis, Raetz pathway, LpxL-LpxM type modules, part of the lipopolysaccharide biosynthesis pathway; the NADH:quinone oxidoreductase, prokaryotes module, part of the oxidative phosphorylation pathway; the tryptophan biosynthesis, chorismate => tryptophan module, part of the phenylalanine, tyrosine and tryptophan biosynthesis pathway, and the 3-Hydroxypropionate bi-cycle module, part of the carbon metabolism and carbon fixation in prokaryotes pathways.

6.4.2 Comparison of microbiome profiles of infected, vaccinated infected, and uninfected animals

6.4.2.1 Diversity indices

The comparison of the alpha and beta diversity estimates for microbiome profiles of animals under different treatments showed that no significant differences were found for alpha diversity. However, BC was significantly different between CHE and VAC at the genera level (0.33 ± 0.03 and 0.26 ±

0.02, respectively, p-value = 0.034), and between UNF and VAC at the microbial genes level (0.22 ± 0.02 and 0.11 ± 0.03, respectively, p-value = 0.015).

6.4.2.2 Pairwise comparison of caecal microbiome profiles of unvaccinated infected and vaccinated infected animals with those of uninfected animals

6.4.2.2.1 Comparison of microbiota profiles

Microbial genera were evaluated for their contribution to the discrimination between groups (in the pairwise PLS-DA comparisons) based on their VIP (i.e., highest VIP corresponds to stronger contribution); microbial genera with VIP ≥ 1 were therefore identified as the most affected by the infection by *O. ostertagi* or by the vaccine against this parasite. The pairwise comparison based on iterative PLS-DA analyses revealed that the first, third, and second PLS-DA models, which included 1194, 204, and 455 microbial genera, led to significant discrimination of CLE, CHE, and VAC, respectively, from UNF in both the PLS-DA and the PCA (Figure 5). In these PLS-DA models discrimination of CLE, CHE, and VAC from UNF, 478, 90, and 167 microbial genera had VIP ≥ 1.

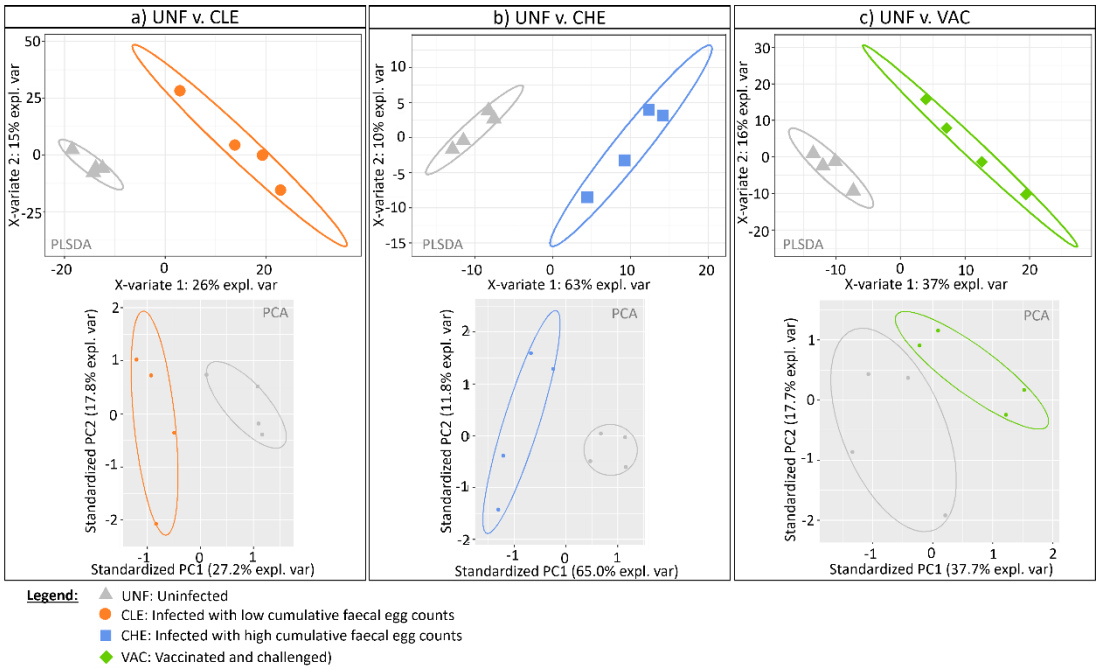


Figure 5. *Partial least squares discriminant analyses (PLS-DA, top figures) and principal component analyses (PCA, bottom figures) plots showing significant discrimination of a) infected animals with low cumulative faecal egg count (CLE), b) infected animals with high cumulative faecal egg count (CHE), and c) vaccinated and challenged animals (VAC), respectively, from uninfected animals (UNF). PLS-DA and PCA models for CLE vs. UNF, CHE vs. UNF, and VAC vs. UNF discriminations included 1194, 204, and 455 genera-level CLR-transformed abundances, respectively.*

The most important microbial genera (i.e., the smallest set of variables able to significantly discriminate the groups, Figure 6) for the CLE vs. UNF discrimination were *Geodermatophilus* and *Pusillimonas* (enriched and depleted, respectively, in CLE), for the CHE vs. UNF discrimination were *Pimelobacter* and *Sulfurospirillum* (enriched and depleted, respectively, in CHE) and for the VAC vs. UNF discrimination were *Glaesserella*, *Oceanimonas*, and *Desulfobacter* (all depleted in VAC).

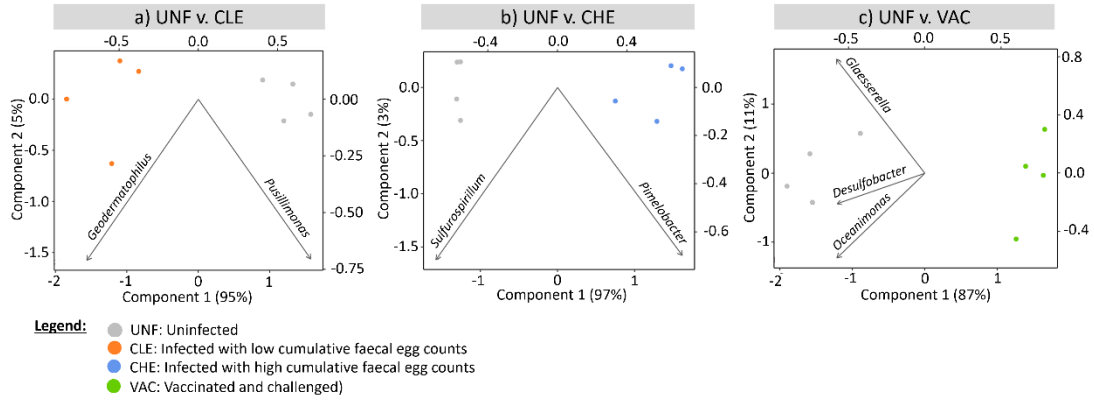


Figure 6. *Smallest set of microbial genera significant discriminating uninfected (UNF) from a) infected with low cumulative faecal egg count (CLE), b) infected with high cumulative faecal egg count (CHE), and c) vaccinated and challenged (VAC). Biplots of partial least squares discriminant analyses using centred logratio-transformed abundances.*

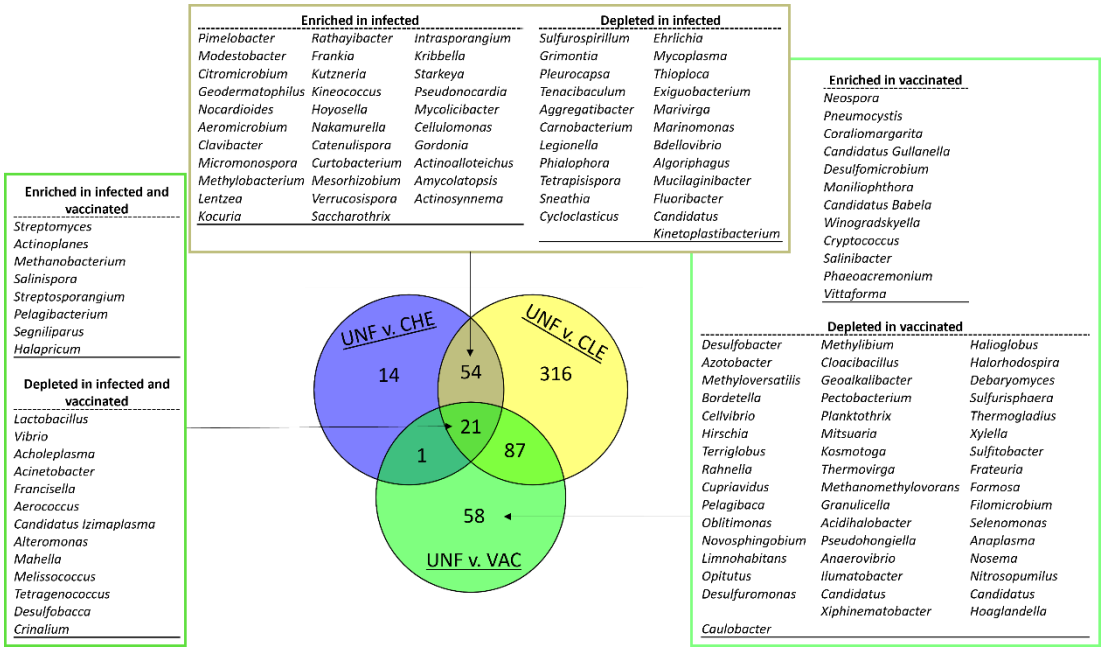


Figure 7. *Microbial genera identified for the discrimination of uninfected (UNF) from infected with low cumulative faecal egg count (CLE), infected with high cumulative faecal egg counts (CHE), and vaccinated and challenged (VAC) animals. "Enriched" and "Depleted" refer to the increased or decreased centre-logratio abundances, respectively, of the microbial genera in CHE, CLE, and/or VAC, in comparison to the abundance in UNF animals.*

Of the microbial genera important for the discrimination of CHE, CLE, and VAC from UNF, 21 were affected in all three groups (vaccinated and both infected

groups with different levels of infection), whereas 54 were exclusively affected in the infected groups. A total of 58 microbial genera were important for the discrimination between VAC and UNF but not for the discrimination between infected (either CLE or CHE) and UNF animals (Figure 7). This suggests that these microbial genera were affected by the vaccine and not by the presence of the nematode itself. Additionally, 14 and 316 microbial genera were differentially affected by the level of infection (important for the discrimination of CHE, and CLE, respectively, from UNF).

6.4.2.2.2 Comparison of microbiome profiles at the functional level

When focusing on microbial genes, the PLS-DA analyses revealed that the PLS-DA models including 6315, 2946, and 911 microbial genes led to the significant separation of CLE, CHE, and VAC, respectively, from UNF (Figure 8). Of these, 3298, 772, and 583 had VIP ≥ 1.

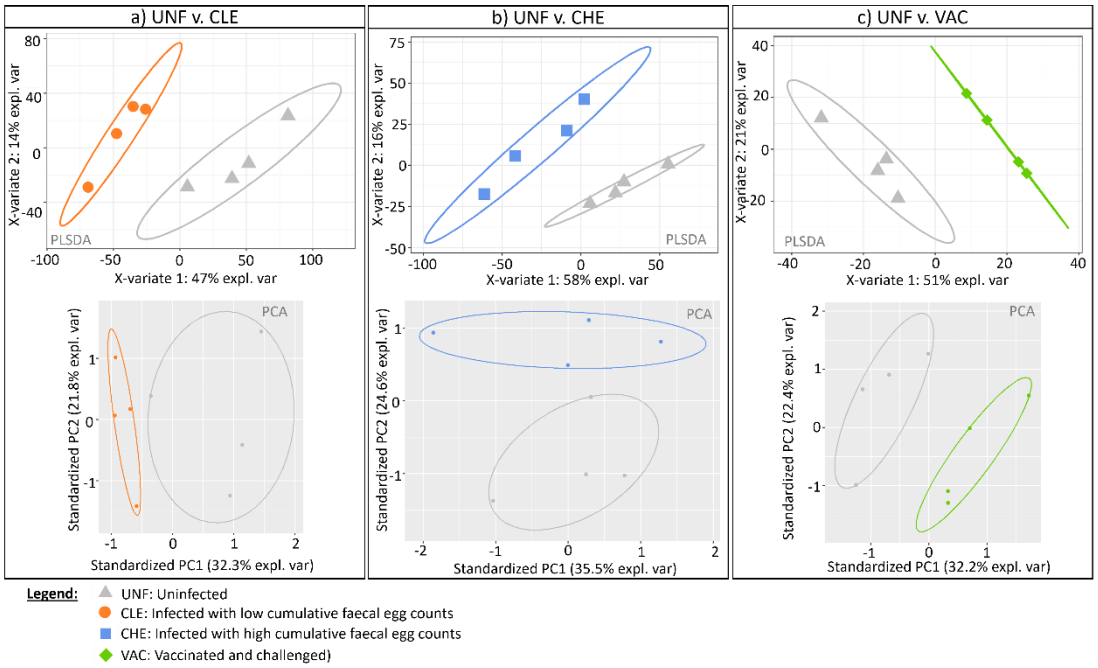


Figure 8. *Partial least squares discriminant analyses (PLS-DA, top figures) and principal component analyses (PCA, bottom figures) plots showing significant discrimination of a) infected animals with low cumulative faecal egg count (CLE), b) infected animals with high cumulative faecal egg count (CHE), and c) vaccinated and challenged animals (VAC), respectively, from uninfected animals (UNF). PLS-DA and PCA models for CLE vs. UNF, CHE*

vs. UNF, and VAC vs. UNF discriminations included 6315, 2946, and 911 microbial genes CLR-transformed abundances.

Most microbial genes enriched in CLE, in comparison to UNF, were associated with (i) energy production and conversion, for example, *aceE* and *sucA* (i.e., pyruvate dehydrogenase E1 component, and 2-oxoglutarate dehydrogenase E1 component, respectively) in the carbon metabolism pathway, and *aldB* and *frdC* (i.e., aldehyde dehydrogenase, and fumarate reductase subunit C, respectively, in the pyruvate metabolism pathway; (ii) carbohydrate transport and metabolism, mostly ABC transporters such as *cebE* and *gtsC* (i.e., cellobiose transport system substrate-binding protein, and glucose/mannose transport system permease protein, respectively) and genes in the starch and sucrose metabolism pathway such as *otsA* and *treY* (i.e. trehalose 6-phosphate synthase, and (1->4)-alpha-D-glucan 1-alpha-D-glucosylmutase); (iii) inorganic ion transport and metabolism, also mostly ABC transporters, e.g., *thiP* and *mntC* (thiamine transport system permease protein, and manganese transport system substrate-binding protein, respectively), but also genes involved in the two-component system pathway, *phoD* and *kdpA* (i.e., alkaline phosphatase D, and potassium-transporting ATPase potassium-binding subunit, respectively) (iv) and the amino acid transport and metabolism, e.g., MAO and E1.5.3.1 (i.e., monoamine oxidase, and sarcosine oxidase, respectively) in the glycine, serine and threonine metabolism pathway, and *dadA* and *phhA* (i.e., D-amino-acid dehydrogenase, and phenylalanine-4-hydroxylase, respectively), in the phenylalanine metabolism pathway.

Most microbial genes depleted in CLE animals were associated with (i) carbohydrate transport and metabolism, mostly involved in carbon metabolism and glycolysis / gluconeogenesis pathway such as *glk* and *pfkA* (i.e., glucokinase, and 6-phosphofructokinase 1, respectively), (ii) amino acid transport and metabolism, including mostly genes involved in biosynthesis of amino acids, such as *hom* and *hisD* (i.e., homoserine dehydrogenase, and histidinol dehydrogenase, respectively), (iii) energy production and

conversion, mostly in the carbon metabolism and pyruvate metabolism, such as *mdh* and *porA* (i.e., malate dehydrogenase, and pyruvate ferredoxin oxidoreductase alpha subunit, respectively), and (iv) cell wall/membrane/envelope biogenesis, mostly in the O-antigen nucleotide sugar biosynthesis pathway, such as UGDH and *rfbD* (i.e., UDPglucose 6-dehydrogenase, and dTDP-4-dehydrorhamnose reductase, respectively) and peptidoglycan biosynthesis pathway, such as *mrcA* and *murA* (i.e., penicillin-binding protein 1A, and UDP-N-acetylglucosamine 1-carboxyvinyltransferase, respectively).

Most microbial genes important for the CHE vs. UNF discrimination were associated with (i) energy production and conversion (8.3% and 5.7% of genes enriched and depleted in CHE, respectively) and carbohydrate transport and metabolism (7% of microbial genes enriched or depleted in CHE). Microbial genes involved in lipid transport and metabolism and amino acid transport and metabolism were mostly enriched in CHE, whereas those involved in cell wall/membrane/envelope biogenesis were mostly depleted in CHE. Genes in the lipid metabolism enriched in CHE were found to be part of the benzoate degradation pathway, e.g., GCDH and *pcaI* (i.e., glutaryl-CoA dehydrogenase, and 3-oxoadipate CoA-transferase, alpha subunit, respectively) and carbon metabolism pathway, e.g., PCCA and *ecm* (i.e. propionyl-CoA carboxylase alpha chain, and ethylmalonyl-CoA mutase, respectively), whereas genes associated with amino acid transport and metabolism mostly belonged to the biosynthesis of amino acids pathway, including *phhA* and GPT (i.e., phenylalanine-4-hydroxylase, and alanine transaminase, respectively). Genes involved in Cell wall/membrane/envelope biogenesis, depleted in infected animals, belonged mostly to the Lipopolysaccharide (LPS) biosynthesis pathway, e.g., *lpxK* and *kdsB* (i.e., tetraacyldisaccharide 4'-kinase, and 3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase), respectively), and to the amino sugar and nucleotide sugar metabolism and O-antigen nucleotide sugar biosynthesis pathways, e.g., *rfbG* and *gmd* (i.e.,

CDP-glucose 4,6-dehydratase, and GDPmannose 4,6-dehydratase, respectively).

The significant discrimination between VAC and UNF was observed in a PLS-DA model using 911 microbial genes, 583 of which had VIP ≥ 1. The 53 microbial genes that were enriched in VAC animals were mostly associated with amino acid transport and metabolism, e.g., GPT (i.e., alanine transaminase) in the alanine, aspartate and glutamate metabolism and arginine biosynthesis pathways and energy production and conversion, e.g., *sdhD* (i.e., succinate dehydrogenase / fumarate reductase, membrane anchor subunit) and *mtmC* (i.e., monomethylamine corrinoid protein), in the carbon metabolism pathway. Microbial genes depleted in VAC were mostly associated with (i) carbohydrate transport and metabolism, including ABC transporters, e.g., *aglK* and *alsB* (i.e., alpha-glucoside transport system ATP-binding protein, and D-allose transport system substrate-binding protein, respectively) and microbial genes participating in pathways such as pentose and glucuronate interconversions, e.g., *uxaB* and *uidA* (i.e., tagaturonate reductase, and beta-glucuronidase, respectively), galactose metabolism, e.g., *galK* (i.e., galactokinase, also in the amino sugar and nucleotide sugar metabolism pathway) and *pfkA* (i.e., 6-phosphofructokinase 1, also in the fructose and mannose metabolism), (ii) amino acid transport and metabolism, most in the biosynthesis of amino acids pathway, e.g., *hisD* and *metH* (i.e., histidinol dehydrogenase, and 5-methyltetrahydrofolate--homocysteine methyltransferase, respectively), (iii) energy production and conversion, e.g., *korC* and *cdhC* (i.e., 2-oxoglutarate ferredoxin oxidoreductase subunit gamma, and acetyl-CoA decarbonylase/synthase, CODH/ACS complex subunit beta, respectively) in the carbon metabolism pathway, LDH and E3.1.2.1 (i.e. L-lactate dehydrogenase, and acetyl-CoA hydrolase, respectively) in the pyruvate metabolism, and ME2 and DLAT (i.e., malate dehydrogenase (oxaloacetate-decarboxylating), and pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase), respectively), in both, and (iv) cell wall/membrane/envelope biogenesis, such as *rfbG* and *gale* (i.e., CDP-

glucose 4,6-dehydratase, and UDP-glucose 4-epimerase, respectively) involved in amino sugar and nucleotide sugar metabolism and O-antigen nucleotide sugar biosynthesis pathways, and *murE* and *mrcB* (i.e., UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase, and penicillin-binding protein 1B, respectively), in the peptidoglycan biosynthesis pathway.

The most important microbial genes for the CLE vs. UNF, CHE vs. UNF and VAC vs. UNF discriminations were, respectively, *rfbG* and E1.1.1.52 (CDP-glucose 4,6-dehydratase, and 3alpha-hydroxycholanate dehydrogenase, respectively, both depleted in CLE), E1.2.7.8 and *colA* (indolepyruvate ferredoxin oxidoreductase, and microbial collagenase, respectively, both enriched in CHE), and *ygjK* and K11777 (putative isomerase, and HAD superfamily phosphatase, respectively, both depleted in VAC) (Figure 9).

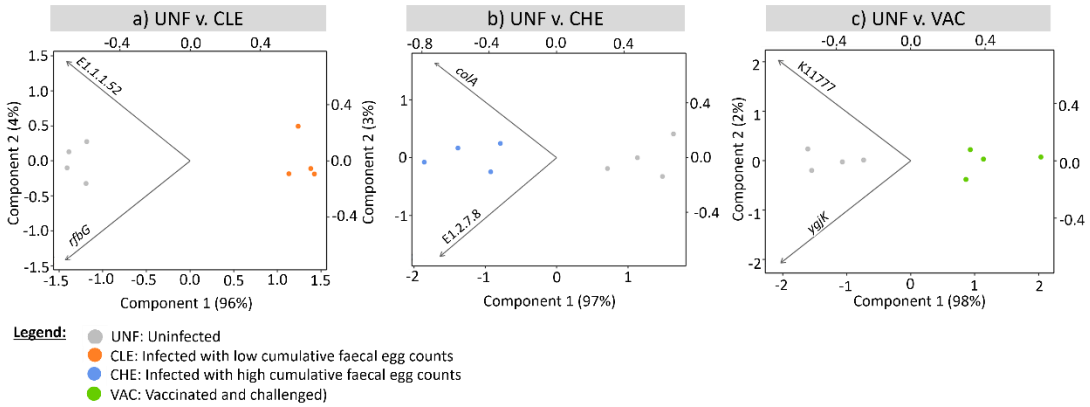


Figure 9. *Smallest set of microbial genes significant discriminating uninfected (UNF) from a) infected with low cumulative faecal egg count (CLE), b) infected with low cumulative faecal egg count (CHE), and c) vaccinated and challenged (VAC); Scores plots of partial least squares discriminant analyses using CLR-transformed abundances.*

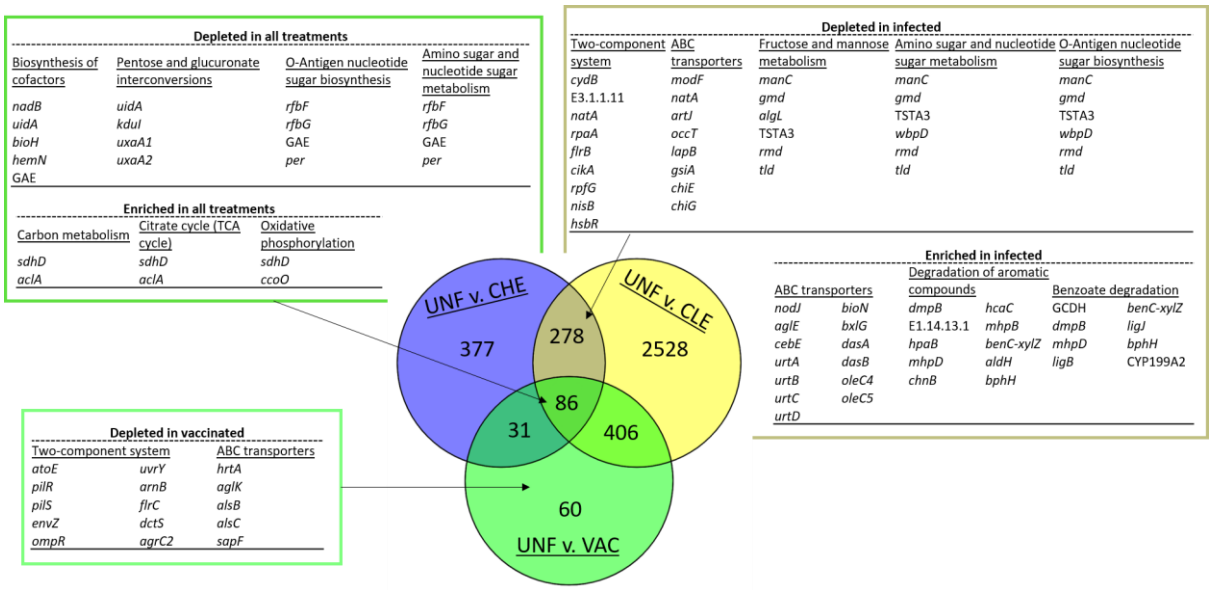


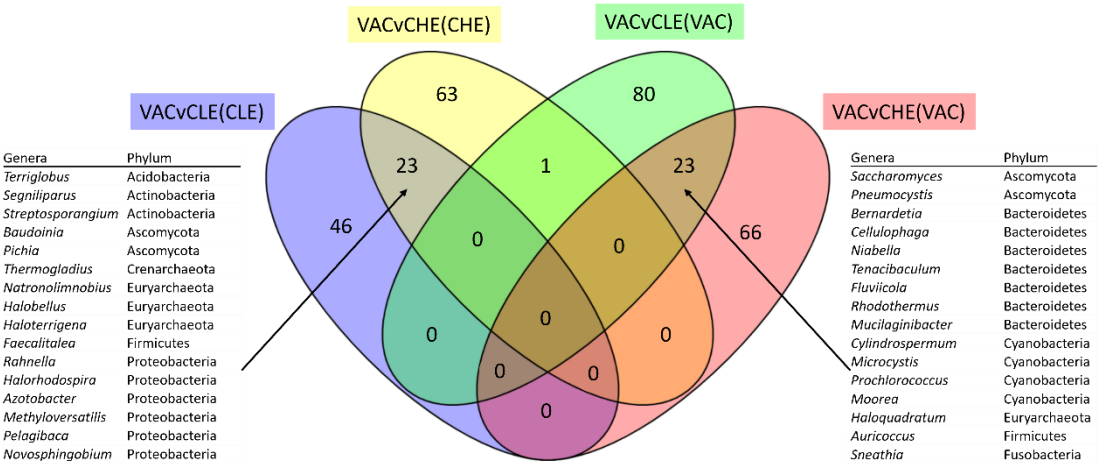
Figure 10. *Microbial genes identified for the discrimination of uninfected (UNF) from infected with low cumulative faecal egg count (CLE), infected with high cumulative faecal egg counts (CHE) and vaccinated and challenged (VAC) animals.*

Similarly to the microbial genera analyses, the analyses based on microbial genes shows that some of them are influenced by the infection independently of the level of infection. Most of these that were depleted in the infected animals were associated with the two-component system pathway, (bacterial communication), amino sugar and nucleotide sugar metabolism, and O-antigen nucleotide sugar biosynthesis, and fructose and mannose metabolism (carbohydrates metabolism), whereas the genes enriched in infected animals were mostly associated with degradation of aromatic compounds, and benzoate degradation. On the other hand, some microbial genes in the two-component system pathway, and some ABC transporters were depleted due to the vaccine (Figure 10).

6.4.2.3 Comparing microbiome profiles of unvaccinated infected with those of vaccinated infected animals

The pairwise comparisons of CHE and CLE microbiome profiles at the genus level with VAC revealed that PLS-DA models using 437 and 425, respectively, led to the significant discrimination of these groups. Of these, 173, and 176

microbial genera had VIP ≥ 1. The results of CHE vs. VAC and CLE vs. VAC revealed that 23 microbial genera of each comparison were observed to be enriched and depleted in the infected (CHE and CLE), in comparison to VAC, respectively (Figure 11).



Genera	Phylum
Terriglobus	Acidobacteria
Segniliparus	Actinobacteria
Streptosporangium	Actinobacteria
Baudoinia	Ascomycota
Pichia	Ascomycota
Thermogladius	Crenarchaeota
Natronolimnobius	Euryarchaeota
Halobellus	Euryarchaeota
Haloterrigena	Euryarchaeota
Faecalitalea	Firmicutes
Rahnella	Proteobacteria
Halorhodospira	Proteobacteria
Azotobacter	Proteobacteria
Methyloversatilis	Proteobacteria
Pelagibaca	Proteobacteria
Novosphingobium	Proteobacteria
Cupriavidus	Proteobacteria
Methylococcus	Proteobacteria
Limnohabitans	Proteobacteria
Desulfuromonas	Proteobacteria
Mitsuaria	Proteobacteria
Thermanaerovibrio	Synergistetes
Kosmotoga	Thermotogae

Genera	Phylum
Saccharomyces	Ascomycota
Pneumocystis	Ascomycota
Bernardetia	Bacteroidetes
Cellulophaga	Bacteroidetes
Niabella	Bacteroidetes
Tenacibaculum	Bacteroidetes
Fluviicola	Bacteroidetes
Rhodothermus	Bacteroidetes
Mucilaginibacter	Bacteroidetes
Cylindrospermum	Cyanobacteria
Microcystis	Cyanobacteria
Prochlorococcus	Cyanobacteria
Moorea	Cyanobacteria
Haloquadratum	Euryarchaeota
Auricoccus	Firmicutes
Sneathia	Fusobacteria
Halobacteriovorax	Proteobacteria
Chania	Proteobacteria
Photorhabdus	Proteobacteria
Idiomarina	Proteobacteria
Candidatus Profftella	Proteobacteria
Candidatus Izimaplasma	Tenericutes
Spiroplasma	Tenericutes

Figure 11. *Microbial genera enriched or depleted in unvaccinated infected in comparison to vaccinated infected animals. The ellipses include microbial genera with VIP ≥ 1 in the partial least squares discriminant analyses comparing vaccinated (VAC) with infected with high cumulative faecal egg count (CHE) or low cumulative faecal egg count (CLE); microbial genera in yellow and blue ellipses were enriched in CHE and CLE, respectively, whereas green and red ellipses include microbial genera enriched in VAC.*

6.5 Discussion

6.5.1 Rumen and caecal microbiome profiles differ significantly

The rumen and caecum have different functions in the gastrointestinal tract of ruminants; whereas the rumen is a fermentation chamber, specialized in the pre-gastric breakdown and digestion of complex polysaccharides including cellulose and hemicellulose, the caecum is the first segment of the large intestine, where previously undigested fibre is fermented by the microbial

communities, and VFAs produced during fermentation are absorbed. Although the caecum contributes substantially less than the rumen to the breakdown and digestion of feed, and consequently to nutrient availability for the host ruminant, the caecum microbiota is still closely associated with the host performance and health, contributing with up to 10% of the dietary energy (Siciliano-Jones and Murphy, 1989; Gressley et al., 2011; O'Hara, 2019); Immig (1996) showed that fermentation in the caecum of ruminants may account for up to 17% of the daily VFA. Fermentation in the caecum may become more important in ruminants when rumen fermentation is hindered (Immig, 1996), due to increased availability of rumen-undigested carbohydrates in the caecum. The comparison of microbiota profiles derived from rumen and caecum samples of uninfected animals showed significantly higher relative abundances of Bacteroidetes and Proteobacteria in the rumen, whereas the caecum had higher relative abundances of Firmicutes and Actinobacteria, which is in agreement with previous studies (Myer et al., 2015; Popova et al., 2017). This reflected the differences in fermentation profiles of the two organs; easily fermentable carbohydrates such as starch, and xylan, are rapidly fermented in the rumen by Bacteroidetes, whereas structural carbohydrates, such as cellulose and hemicellulose, may escape the ruminal fermentation and reach the caecum, where they are utilized by Firmicutes (Siciliano-Jones and Murphy, 1989; Gressley et al., 2011; Min et al., 2019).

The difference between microbiota profiles of rumen and caecum was further highlighted in the stacked bar charts obtained from the relative abundances of the most abundant microbial genera in uninfected animals, where we observed the rumen microbiota to be dominated by *Prevotella*, *Methanobrevibacter*, *Succiniclasticum*, *Ruminococcus*, and *Fibrobacter*, whereas the caecum microbiota was dominated by *Ruminococcus* and *Methanobrevibacter*, and had relatively lower abundances of *Prevotella, Succiniclasticum*, and *Fibrobacter*. *Prevotella* is a predominant member of the rumen core microbiota (Holman and Gzyl, 2019) thought to be mainly involved in the non-cellulose plant fibre degradation, such as hemicellulose and glycans, and it was

previously found to be enriched in foregut samples in comparison to those from the small and large intestine in dairy cattle (Mao et al., 2015); *Prevotella* was significantly more abundant in the rumen, and this is likely due to their role in the ruminal degradation of starch and proteins as well as uptake and fermentation of peptides (Myer et al., 2015).

For the diversity at the family and genus level, we observed significantly higher unadjusted and adjusted Shannon indices in the caecum, indicating that this sample type is characterized by a more even distribution of the microbial communities than the rumen. Mao et al. (2015) found, in the gastrointestinal tract microbiota of dairy cattle, significantly higher evenness of OTUs in the rumen than in the caecum (Shannon indices of 6.26 and 3.13, respectively), which disagrees with our results. This difference may be due to biological differences or due to the bioinformatics pipelines applied to identify taxonomic groups; whereas Mao et al. (2015) applied QIIME to obtain the taxonomic profiles based on 16S rRNA gene amplicons using clustering at 97% similarity, we applied the Kraken approach on sequence reads obtained from whole metagenomic shotgun sequencing using as reference databases the Hungate 1000 collection (Seshadri et al., 2018) and MAGs generated from beef rumen samples (Stewart et al., 2018). Whole metagenome sequencing based methods have been previously suggested to provide more accurate results than 16S rRNA-based methods (Jovel et al., 2016).

For diversity at microbial genes level, the observed richness showed a tendency to be higher in the caecum. We observed significantly higher evenness in the rumen, which could be associated with the microbial fermentation profiles of these organs. Whereas the feed that enters the rumen contains high diversity of substrates, from easily fermentable compounds such as starch, to difficult to break-down complex polysaccharides like cellulose, the caecum will most likely only receive some of the complex polysaccharides that were previously not digested in the rumen, leading therefore to the growth of microbial taxa that carry in their genomes the microbial genes associated with

cellulose and hemicellulose degradation, and not so much those associated with starch degradation.

The analyses of the pathways in which the most important microbial genes for the discrimination between caecum and rumen microbiome profiles are included revealed 1018 and 798 microbial genes enriched, respectively, in the caecum and rumen microbiome profiles. The analyses of the pathways and modules in which these microbial genes participate suggested that the ruminal and caecal microbial gene profiles differ significantly in their functional potential. The rumen was enriched in microbial genes associated with environmental sensing, and cell wall structures, whereas the caecum was richer in microbial genes associated with carbon metabolism, and methanogenesis.

6.5.2 Infection by the parasitic nematode *Ostertagia ostertagi*, and vaccination against this parasite, affects the caecum microbiota

The significant discrimination between CLE and UNF was observed when including 1194 microbial genera in the explanatory dataset, of which 478 had VIP ≥ 1, 156 enriched and 322 depleted in CLE, showing that the presence of *O. ostertagi* has an impact on the overall structure of the caecal microbiome. In contrast, the significant discrimination between CHE and UNF included 204 microbial genera. This lower number of genera suggests that the influence of *O. ostertagi* parasitism in the caecal microbiome of more susceptible animals (i.e., with higher cFEC) affected microbial communities in a more specific manner than in more resilient animals (i.e., with low cFEC). Of these 204 microbial genera, 90 had VIP ≥ 1, of which 47 and 43 were enriched and depleted, respectively, in CHE.

Within the 156 genera whose abundance was enriched in CLE vs. UNF, 69 (44%) belong to the Actinobacteria and 24 (15%) to Proteobacteria phyla, whereas out of the 322 microbial genera depleted in CLE, 99 (31%) were Proteobacteria and 68 (21%) were Firmicutes. The most important microbial

genera enriched in CLE were mostly from the class Actinomycetia, which are high G+C, Gram-positive bacteria that most often inhabit the soil such as *Geodermatophilus* (Gordon and Perrin, 1971) and *Rhodococcus* (Willey et al., 2020a). *Kineococcus* is also an Actinomycetia and was previously found in pigs infected with *Trichuris suis* (Li et al., 2012). Additionally, we observed enrichment of opportunistic pathogens such as *Babesia* and *Toxoplasma*, in agreement with the enrichment of opportunistic pathogens, such as *Streptomyces* and *Tsukamurella*, in the rumen of infected in comparison to uninfected animals, reported in chapter 5.

The enrichment of microorganisms in the caecum of infected animals that typically inhabit the soil, together with increased abundances of opportunistic pathogens, may be associated with the putatively increased pH in the abomasum, since a low pH in the abomasum is recognized as a barrier, protecting the rest of the gastrointestinal tract (Constable et al., 2006).

Most microbial genera enriched in CHE belong to phylum Actinobacteria (mostly to families *Pseudonocardiaceae* and *Microbacteriaceae*).
Pseudonocardiaceae was previously reported to have increased relative abundance in the gut microbiota of mice with cystic fibrosis (Lynch et al., 2013), and *Microbacteriaceae* has been found to be enriched in chronically diseased sigmoid colon tissue, in comparison to adjacent tissues, in humans with chronic, recurrent diverticulitis (Schieffer et al., 2017). Out of the 47 microbial genera with VIP ≥ 1 and enriched in CHE in this comparison, 40 were also identified in the CLE vs. UNF comparison (also with VIP ≥ 1 and enriched in the infected group), mostly belonging to the class Actinomycetia and some Alphaproteobacteria, including e.g., the methylotrophic *Methylobacterium* (Patt et al., 1976). The methanogenic archaea *Methanobacterium* was enriched, whereas the nitrate-reducing *Veillonella* was depleted in infected animals. Interestingly, Iwamoto et al. (2002) reported a negative association between nitrate-reducing organisms and methanogenesis in the rumen; whereas nitrate-reducing organisms are tolerant to the nitrite toxicity,

methanogens are not. Additionally, *Marinomonas*, which was previously reported as having negative effect on methanogenesis in the rumen (Auffret et al., 2018) was here depleted in infected animals.

A total of 40 microbial genera were important for the CHE vs. UNF and CLE vs. UNF discriminations, and were enriched in both infected groups, in comparison to UNF. These included 33 actinomycetes, e.g., *Cellulomonas*, *Frankia*, and *Streptomyces*, which can produce a variety of secondary metabolites, including antibiotics. Since WGS methods do not distinguish between spores and live bacterial organisms, and some of these microbial genera, for example, *Intrasporangium* and *Micromonospora*, are spore-forming bacteria, it is possible that they are present in the caecum of infected animals in their dormant form.

Most microbial genera important for the CLE vs. UNF discrimination, depleted in CLE, belonged to classes Gammaproteobacteria (12 out of 43 – 28%) and Bacilli (8 out of 43 – 19%). Lactobacilli *Aerococcus*, *Carnobacterium*, *Tetragenococcus*, *Melissococcus*, and *Lactobacillus* were depleted in infected animals. In a murine study in which mice were infected with the intestinal nematode *Heligmosomoides polygyrus*, Reynolds et al. (2016) showed that *Lactobacillus* was positively correlated with the infection level in mice highly susceptible to infection, and that the correlation between *Lactobacillus* and infection level was negative in more resistant mice. In agreement, in our study, in which we infected calves at subclinical levels, the lowest abundance of *Lactobacillus* was observed in CHE animals, whereas the highest abundance was observed in UNF.

Filifactor was the genus with the highest VIP in the CLE vs. UNF comparison and was depleted in CLE animals. This genus belongs to the class *Clostridia*, phylum *Firmicutes* and two species have previously been described, *F. villosus* (the type species) and *F. Alocis*. *F. villosus* was previously isolated from subcutaneous abscesses in cats and described as *Clostridium villosum* by Love et al. (1979) as an obligately anaerobic, rod-shaped, spore-forming, non-

motile organism which is Gram-positive in its early stages of growth and Gram-negative after 18 to 24 h, and was later reclassified (Collins et al., 1994). Strains of the species were shown to grow well in a culture medium of cooked meat plus peptic digest of meat (CMM) broth supplemented with glucose, cellobiose, maltose and starch (CMC), and it was shown to produce more butyrate in culture mediums containing pyruvic acid. *F. alocis* is a non-spore-forming, Gram-negative, obligately anaerobic rod that produces butyrate and acetate in a peptone-yeast extract-glucose medium (PYG). Both species had their growth stimulated by the addition of animal-origin serums (i.e., rabbit in *F. alocis* and horse in *F. villosum*) to the mediums. This genus was later identified in the rumen epimural bacterial community of forage-fed heifers of beef cattle, while it was absent from their mixed-forage and high-grain fed counterparts as well as from the rumen while in induced acidotic challenge (Petri et al., 2013) and more recently, it was detected in the microbiota of the rumen liquid digesta (where it was significantly more abundant than in the solid phase) of heifers euthanized on the 7th day after vaginal delivery but it was not detected in the rumen microbiota of heifers euthanized from day 14 onwards (O'Hara, 2019). Additionally, *Agarivorans*, *Pragia*, *Gallibacterium*, and *Grimontia* (Gammaproteobacteria) also had high VIPs in this comparison and were depleted in CLE.

Thirty-five microbial genera were important in both CHE vs. UNF and CLE vs. UNF discriminations, being depleted in infected animals, including several groups with pathogenic potential, belonging to Gammaproteobacteria, e.g., *Acinetobacter*, *Legionella*, and *Vibrio*, and Bacilli, e.g., *Aerococcus*, and *Melissococcus*. The depletion of opportunistic pathogens in the caecum of infected in comparison to uninfected animals was unexpected, particularly considering the enrichment of microbial taxa with pathogenic potential in the rumen, reported in chapter 5. Additionally, we observed depletion of beneficial lactic acid producers in both groups of infected animals, including, for example, *Lactobacillus*, *Tetragenococcus*, and *Carnobacterium*.

Four hundred and fifty-five microbial genera were included in the PLS-DA that led to the significant discrimination of UNF from VAC, of which 167 had VIP ≥ 1. Of these microbial genera, 39 and 128 were enriched and depleted, respectively, in VAC, in comparison to UNF. Eight and seven out of the 39 microbial genera enriched in VAC (21% and 18%, respectively) belonged to the phyla Actinobacteria and Proteobacteria, respectively, whereas Proteobacteria (69 out of 128 = 54%) and Firmicutes (20 out of 128 = 16%) were the most dominant phylum among the microbial genera depleted in VAC. All Actinobacteria enriched in VAC were Actinomycetia; *Streptosporangium* was previously identified in the rumen microbiota of dairy cattle whose milk had low abundance of saturated fatty acids (Stergiadis et al., 2021), and the Gram-positive *Micrococcus* is an ubiquitous anaerobe (found in soil, water, and vegetation, as well as on the skin of warm blooded animals), and are often reported as opportunistic pathogens, belonging to the same family as *S. aureus* and *S. haemolyticus* (*Micrococcaceae*), which are associated with mastitis in dairy cattle (Nuñez, 2014). This genus was previously identified as part of the intestinal flora of cattle fed high-roughage rations (Maki and Picard, 1965).

Within the Proteobacteria enriched in VAC, we identified *Methyloceanibacter*, a methylotroph previously shown to negatively correlate to methane production in beef cattle (Auffret et al., 2018). The methanogen *Methanobacterium* was also enriched, whereas methanogen *Methanomethylovorans* and methylotroph *Methyloversatilis* were depleted in VAC.

Desulfovibrio is a Gram-negative anaerobic acetate-producer in the *Desulfovibrionales* order, which is a diverse group of sulphur-reducing bacteria (Willey et al., 2020b) and has previously been shown to be associated with sulphate reduction in the sheep rumen (Howard and Hungate, 1976). D*esulfomicrobium* is one of the most often occurring sulphate reducing bacteria in GIT of humans and animals (Dordević et al., 2021). The enrichment of these microbial genera in VAC in comparison to UNF may be either due to

increased sulphate availability in the caecum of VAC or to improved use of sulphate in the caecum of UNF. Sulphate has been suggested as an effective hydrogen sink, leading to decreased methane production (but increased fermentation rates) by rumen microorganisms; additionally, sulphate was suggested to increase reduction of nitrite to ammonia (van Zijderveld et al., 2010). For example, *Desulfovibrio* was reported as enriched in the jejunum of steers with higher growth rates (Freetly et al., 2020), it was twice more abundant in high methane emitting cattle than in the low-emitting counterparts (Wallace et al., 2015), and it was enriched in the colon microbiota of pigs infected with the nematode *Trichuris suis* (Li et al., 2012). However, we found other sulfate-reducing genera to be depleted in vaccinated animals, including *Desulfobacter* and *Desulfocapsa* (order *Desulfobacterales*) and *Geoalkalibacter* and *Desulfuromonas* (order *Desulfuromonadales*) (Willey et al., 2020b).

We also observed enrichment of several fungi in the caecum of vaccinated animals, including Ascomycota, e.g. *Pneumocystis*, a potential bovine pathogen (Settries and Henriksen, 1989), Basidiomycota, e.g. *Malassezia*, an opportunistic pathogen in animals (Summerbell, 2004), *Cryptococcus*, associated with mastitis in cattle (Summerbell, 2004), and *Moniliophthora*, correlated with methane production (Martínez-Álvaro et al., 2021) and Microsporidia, e.g., *Vittaforma*, a human intestinal and urinary pathogen (Mathis et al., 2005). The enrichment of opportunistic pathogenic fungi belonging to phyla Ascomycota and Basidiomycota was also observed in the rumen of vaccinated animals (chapter 5).

A total of 128 microbial genera were depleted in vaccinated animals, most of them belonging to the phylum Proteobacteria. For example, *Gluconobacter* is an Alphaproteobacteria previously associated with low methane production in beef cattle (Martínez-Álvaro et al., 2020), and *Cronobacter* is a Gammaproteobacteria previously associated with lower methane production in beef cattle (Auffret et al., 2018).

Bordetella (Betaproteobacteria) is a widely recognized respiratory pathogen; it is an aerobic Gram-negative proteobacterium, which requires organic sulfur and nitrogen for growth (Willey et al., 2020b). However, it has been identified in the rumen of young calves (Malmuthuge et al., 2019), which, like the caecum, is typically characterized as an anaerobic environment. The murine respiratory tract microbiota has been shown to inhibit the growth of *B. pertussis* (Weyrich et al., 2014), and in the present work *Bordetella* was found to be significantly depleted in vaccinated and enriched in infected (although with VIP < 1), in comparison to UNF. It is plausible to assume that the degree of colonization by a pathogen such as *Bordetella* is dependent on the fitness of the microbiota in the caecum, which is expected to be hindered in infected animals and strengthened in vaccinated animals, either due to direct effects of these factors on the microbiota, or through the host's immune system influence on the microbiota.

Enterobacter are facultative anaerobes, butanediol and lactose fermenters that produce butanediol, ethanol and CO_2 (Willey et al., 2020b). Species of this genus have been identified in the rumen fluid and attached to the rumen wall (Mitsumori et al., 2002). Members of the same family (i.e. *Enterobacteriaceae*) were identified in the gastrointestinal track of cattle, where they were enriched in the small intestine, caecum and colon in comparison to the foregut, and one Enterobacteriaceae OTU (OTU-3825) was the most abundant in the duodenum microbiota (Mao et al., 2015). *Enterobacteriaceae* has also been associated with nematode infections; Rausch et al. (2013) reported higher abundances of Gram-negative *Gammaproteobacteria*/*Enterobacteriaceae* in the caecum and colon of mice infected with the nematode *Heligmosomoides polygyrus bakeri* in comparison to their uninfected counterparts, and suggested that the increased growth of *Enterobacteriaceae* in the intestinal lumen of infected animals could be due to increased glucose availability, which is not efficiently absorbed in infected animals (Olaogun and Lasisi, 2015). Reynolds et al. (2016) showed that susceptible C57BL/6 mice infected with the

nematode *Heligmosomoides polygyru* had significantly higher levels of *Enterobacteriaceae* in the duodenum, in comparison to naïve animals.

6.5.3 Influence of the presence of *O. ostertagi* and the vaccine against this abomasal parasite on the caecum microbial genes

Several microbial genes were found to be altered in CHE, CLE, and VAC, in comparison to UNF. For example, *uidA* (i.e., beta-glucuronidase) is a microbial gene associated with carbohydrate metabolism, participating in the biosynthesis of cofactors and the pentose and glucuronate interconversions, and was depleted in CHE, CLE, and VAC, in comparison to UNF; *uidA* was previously shown to be involved in the degradation of xylan (Lee et al., 2012), and it was identified as a rumen microbial gene biomarker for decreased feed conversion efficiency in beef cattle (Roehe et al., 2016; Lima et al., 2019). Furthermore, microbial genes *rfbF* and *rfbG* (i.e., glucose-1-phosphate cytidylyltransferase, and CDP-glucose 4,6-dehydratase, respectively), depleted in CHE, CLE, and VAC are involved in cell wall biosynthesis (in the O-antigen nucleotide sugar biosynthesis and the amino sugar and nucleotide sugar nucleotide pathways). These genes are part of the *rfb* operon, and are associated with microbial lipopolysaccharide (LPS) production, which together with the O-antigen (both in the outer membrane of Gram-negative bacteria), is important for host colonization and niche adaptation, and plays a part in protection of microbial communities from host immune responses (Reeves, 1995). Additionally, these microbial genes were previously was associated with appetite in beef cattle, showing a positive correlation with daily feed intake (Lima et al., 2019).

Some microbial genes depleted in CHE and CLE (but not in VAC), in comparison to UNF, were also associated with the O-antigen nucleotide sugar biosynthesis (e.g., *manC*, *gmd*). The O-antigen is an important component of the cell wall of Gram-negative bacteria, involved in the host immune response (Samuel and Reeves, 2003). Microbial genes depleted in CHE and CLE were also found to be associated with environmental sensing, belonging to the two-

component system, and included ABC transporters. For example, *artJ*, and *modF* (arginine transport system substrate-binding protein, and molybdate transport system ATP-binding protein, respectively, ABC transporters), were previously shown to be uniquely and highly expressed by *Salmonella Typhimurium* in a magnesium-depleted medium designed to mimic the macrophage phagosomal environment (Adkins et al., 2006). These results suggest that the influence of the parasite on the caecal microbiome may happen via host immune system response.

Microbial genes depleted in CHE and CLE included also e.g., the pectinesterase EC:3.1.1.11, which catalyses the de-esterification of pectin into pectate and methanol, with release of hydrogen (AmiGO 2, 2021; EMBL 2021, 2021). Since methanogenesis is the main hydrogen sink in the bovine GIT, and methanogens can utilize methanol as electron acceptor (Castillo-González et al., 2014), the depletion of this gene is a possible indicator of hindered methane production in infected animals, such as reported in the rumen of infected sheep (Fox et al., 2018).

Microbial genes associated with the uptake of urea from the environment (Valladares et al., 2002), e.g., *urtA*, *urtB*, *urtC*, and *urtD* (i.e., urea transport system substrate-binding protein, urea transport system permease protein, urea transport system permease protein, and urea transport system ATP-binding protein, respectively) were enriched in CHE and CLE.

Additionally, microbial genes *hcaC*, *mhpB*, *mhpD* (i.e., 3-phenylpropionate/trans-cinnamate dioxygenase ferredoxin component, 2,3-dihydroxyphenylpropionate 1,2-dioxygenase, and 2-keto-4-pentenoate hydratase, respectively) were enriched in CHE and CLE, in comparison to UNF. These genes participate in the degradation of aromatic compounds pathway, which has been suggested as involved in the virulence of *Escherichia coli* B2 strains (Touchon et al., 2009).

6.5.4 Comparison of microbiome profiles of unvaccinated infected with vaccinated infected animals

Microbial genes depleted in VAC, in comparison to UNF, included the two-component system *envZ*/*ompR* (i.e., two-component system, OmpR family, osmolarity sensor histidine kinase *EnvZ*, and phosphate regulon response regulator *OmpR*, respectively), associated with the regulation of the expression of outer membrane proteins in response to changes in the environmental osmolarity (Kenney and Anand, 2020), and microbial gene *uvrY* (i.e., two-component system, NarL family, invasion response regulator *UvrY*), which is part of the two component system *barA*/*uvrY*, which controls carbon metabolism (Pernestig et al., 2003).

The pairwise comparisons CHE vs. VAC and CLE vs. VAC revealed depletion of several microbial genes in the caecal microbiome of infected animals that are involved in two-component systems, e.g., the sensor kinase *cheA* (two-component system, chemotaxis family), which was previously identified as associated with improved feed conversion efficiency in beef cattle (Lima et al., 2019); ABC transporters, e.g., *natA* and *natB* (i.e., sodium transport system permease protein, and sodium transport system ATP-binding protein, respectively) are associated with the uptake of amino acids (Montesinos et al., 1997; Hong et al., 2017); and flagellar assembly, e.g., *fliA*, *fliC*, *fliE*, and *flip* (i.e., RNA polymerase sigma factor *FliA*, flagellin, flagellar hook-basal body complex protein *FliE*, and flagellar biosynthesis protein *FliP*, respectively). Microbial genes *nisB* and *nisC* (i.e., both lantibiotic biosynthesis proteins), are involved in the production of the bacteriocin nisin, which inhibits most Gram-positive bacteria (Li and O'Sullivan, 2006).

6.6 Conclusions

Our comparison between caecal and ruminal microbiome profiles showed that the differences observed in the microbiome reflect the functional differences between these organs. Since the rumen microbiota receives feed as it is ingested, with more diverse substrates available for microbial fermentation, its

biochemical networks sustained by the resident microbial communities are more diverse, as suggested by the significantly higher adjusted Shannon index. In contrast, the caecum will receive only the substrates that escape ruminal fermentation, richer in fibrous matter, and less rapidly fermentable substrates like starch, which is reflected in the decreased abundance of Bacteroidetes and increased abundance of Firmicutes. Furthermore, differences were observed at the functional level of analyses; whereas the caecum was richer in microbial genes associated with, for example, carbon metabolism and methanogenesis, the rumen microbiome was enriched in microbial genes with functions associated with cell wall structures and environmental sensing.

Overall, the comparisons of microbiome profiles of treatment groups as comparisons of infected showing high cumulative faecal egg count (CHE), infected showing low cumulative faecal egg count (CLE), and vaccinated and challenged (VAC) with uninfected (UNF) animals indicate that the infection by the abomasal nematode *O. ostertagi* even at subclinical levels, leads to a marked dysbiosis in the caecal microbiome profiles of dairy cattle.

The pairwise comparison of infected with uninfected animals revealed enrichment of several actinomycetes in the infected groups, e.g., *Cellulomonas*. In addition, enrichment of microbial genes associated with uptake of urea, such as *urtA*, and *urtB*, important for microbial protein metabolism, and with degradation of aromatic compounds, such as *hcaC*, and *mhpB*, associated with microbial virulence, was observed in the caecum of infected, in comparison to uninfected animals.

The caecum of infected, in comparison to uninfected animals, was depleted of several opportunistic pathogenic Gammaproteobacteria, such as *Acinetobacter*, and Bacilli, such as *Aerococcus*. This result was unexpected, particularly considering the enrichment of opportunistic pathogens such as *Streptomyces, and Tsukamurella* in the rumen microbiome, reported in chapter 5 of this thesis. Favourable lactic acid producers, including *Lactobacillus* were

also depleted in infected, in comparison to uninfected animals. Additionally, we observed depletion of microbial genes with functions in carbohydrates transport and metabolism, and environmental sensing, including ABC transporters and microbial genes included in the O-antigen nucleotide sugar biosynthesis, and the two-component system pathways, in infected in comparison to uninfected animals.

We have demonstrated that the native vaccine against *O. ostertagi* caused alterations to the microbiome profiles in the caecum, at both taxonomic and functional levels. The caecum of vaccinated animals was enriched in several fungi with pathogenic potential including genera of the phyla Ascomycota (e.g., *Pneumocystis*) and Basidiomycota (e.g., *Malassezia*), and depleted of Proteobacteria, including *Cronobacter*. These results agreed with those reported in chapter 5 of this thesis, which also showed enrichment of microbial fungi with pathogenic potential in the rumen of vaccinated animals. We also observed depletion of microbial genes associated with environmental sensing such as those included in the two-component system pathway in the caecum of unvaccinated infected in comparison to vaccinated infected animals.

The results of this study highlight that an infection by abomasal nematode *O. ostertagi* changes the microbial profiles and their functions in the caecum which should be considered in the development of vaccines or used for the development of pro- and prebiotics to reduce at least the clinical signs of ostertagiasis and its impact on loss of performance.

6.7 References

Adkins, J. N., Mottaz, H. M., Norbeck, A. D., Gustin, J. K., Rue, J., Clauss, T. R. W., et al. (2006). Analysis of the Salmonella typhimurium proteome through environmental response toward infectious conditions. *Mol. Cell. Proteomics* 5, 1450–1461. doi:10.1074/mcp.M600139-MCP200.

AmiGO 2 (2021). Pectinesterase activity. Available at: http://amigo.geneontology.org/amigo/term/GO:0030599.

Auffret, M. D., Stewart, R., Dewhurst, R. J., Duthie, C.-A., Rooke, J. A., Wallace, R. J., et al. (2018). Identification, comparison, and validation of robust rumen microbial biomarkers for methane emissions using diverse Bos taurus breeds and basal diets. *Front. Microbiol.* 8, 1–15. doi:10.3389/fmicb.2017.02642.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–63.

Bueno, L., Dakkak, A., and Fioramonti, J. (1982). Gastro-duodenal motor and transit disturbances associated with Haemonchus Contortus infection in sheep. *Parasitology* 84, 367–374. doi:10.1017/S0031182000044905.

Cao, K.-A. Le, Rohart, F., Gonzalez, I., Dejean, S., Abadi, A., Gautier, B., et al. (2020). Package 'mixOmics.'

Castillo-González, A. R., Burrola-Barraza, M. E., Domínguez-Viveros, J., and Chávez-Martínez, A. (2014). Rumen microorganisms and fermentation. *Arch. Med. Vet.* 46, 349–361. doi:10.4067/S0301-732X2014000300003.

Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., et al. (1994). The phylogeny of the genus Clostridium: Proposal of five new genera and eleven new species combinations. *Int. J. Syst. Bacteriol.* 44, 812–826. doi:10.1099/00207713-44-4-812.

Constable, P. D., Wittek, T., Ahmed, A. F., Marshall, T. S., Sen, I., and Nouri, M. (2006). Abomasal pH and emptying rate in the calf and dairy cow and effect of commonly adminstered therapeutic agents. *XXIV World Buiatrics Congr.*, 54–68.

de Oliveira, M. N. V., Jewell, K. A., Freitas, F. S., Benjamin, L. A., Tótola, M. R., Borges, A. C., et al. (2013). Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. *Vet. Microbiol.* 164, 307–314. doi:10.1016/j.vetmic.2013.02.013.

Difford, G. F., Plichta, D. R., Løvendahl, P., Lassen, J., Noel, S. J., Højberg, O., et al. (2018). Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genet.* 14, 1–22. doi:10.1371/journal.pgen.1007580.

Dordević, D., Jančíková, S., Vítězová, M., and Kushkevych, I. (2021). Hydrogen sulfide toxicity in the gut environment: Meta-analysis of sulfate-reducing and lactic acid bacteria in inflammatory processes. *J. Adv. Res.* 27, 55–69. doi:10.1016/j.jare.2020.03.003.

Edmonds, M. D., Johnson, E. G., and Edmonds, J. D. (2010). Anthelmintic resistance of Ostertagia ostertagi and Cooperia oncophora to macrocyclic lactones in cattle from the western United States. *Vet. Parasitol.* 170, 224–229. doi:10.1016/j.vetpar.2010.02.036.

EMBL 2021 (2021). EC 3.1.1.11 - Pectinesterase. Available at: https://www.ebi.ac.uk/intenz/query?cmd=SearchEC&ec=3.1.1.11.

Fox, M. T., Gerrelli, D., Pitt, S. R., Jacobs, D. E., Gill, M., and Gale, D. L. (1989). Ostertagia ostertagi infection in the calf: effects of a trickle challenge on appetite, digestibility, rate of passage of digesta and liveweight gain. *Res. Vet. Sci.* 47, 294–298. doi:10.1016/s0034-5288(18)31249-9.

Fox, M. T., Gerrelli, D., Pitt, S. R., Jacobs, D. E., Hart, I. C., and Simmonds, A. D. (1987). Endocrine effects of a single infection with Ostertagia ostertagi in the calf. *Int. J. Parasitol.* 17, 1181–1185. doi:10.1016/0020-7519(87)90170-6.

Fox, M. T., Reynolds, G. W., Scott, I., Simcock, D. C., and Simpson, H. V. (2006). Vagal and splanchnic afferent nerves are not essential for anorexia associated with abomasal parasitism in sheep. *Vet. Parasitol.* 135, 287–295. doi:10.1016/j.vetpar.2005.10.015.

Fox, M. T., Uche, U. E., Vaillant, C., Ganabadi, S., and Calam, J. (2002). Effects of Ostertagia ostertagi and omeprazole treatment on feed intake and gastrin-related responses in the calf. *Vet. Parasitol.* 105, 285–301. doi:10.1016/S0304-4017(02)00026-2.

Fox, N. J., Smith, L. A., Houdijk, J. G. M., Athanasiadou, S., and Hutchings, M. R. (2018). Ubiquitous parasites drive a 33% increase in methane yield from livestock. *Int. J. Parasitol.* 48, 1017–1021. doi:10.1016/j.ijpara.2018.06.001.

Freetly, H. C., Dickey, A., Lindholm-Perry, A. K., Thallman, R. M., Keele, J. W., Foote, A. P., et al. (2020). Digestive tract microbiota of beef cattle that differed in feed efficiency. *J. Anim. Sci.* 98, 1–16. doi:10.1093/jas/skaa008.

Geerts, S., Brandt, J., Kumar, V., and Biesemans, L. (1987). Suspected resistance of Ostertagia ostertagi in cattle to levamisole. *Vet. Parasitol.* 23, 77–82. doi:10.1016/0304-4017(87)90026-4.

Gordon, M. A., and Perrin, U. (1971). Pathogenicity of Dermatophilus and Geodermatophilus. *Infect. Immun.* 4, 29–33. doi:10.1128/iai.4.1.29-33.1971.

Greenacre, M. (2018). *Compositional Data Analysis in Practice.* , ed. C. & H. / C. Press.

Gressley, T. F., Hall, M. B., and Armentano, L. E. (2011). Ruminant nutrition symposium: Productivity, digestion, and health responses to hindgut acidosis in ruminants. *J. Anim. Sci.* 89, 1120–1130. doi:10.2527/jas.2010-3460.

Guan, L. L., Nkrumah, J. D., Basarab, J. A., and Moore, S. S. (2008). Linkage of microbial ecology to phenotype: Correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol. Lett.* 288, 85–91. doi:10.1111/j.1574-6968.2008.01343.x.

Holman, D. B., and Gzyl, K. E. (2019). A meta-analysis of the bovine gastrointestinal tract microbiota. *FEMS Microbiol. Ecol.* 95, 1–9. doi:10.1093/femsec/fiz072.

Hong, H., Cai, Y., Zhang, S., Ding, H., Wang, H., and Han, A. (2017). Molecular basis of substrate specific acetylation by N-terminal acetyltransferase NatB. *Structure* 25, 641-649.e3. doi:10.1016/j.str.2017.03.003.

Howard, B. H., and Hungate, R. E. (1976). Desulfovibrio of the sheep rumen. *Appl. Environ. Microbiol.* 32, 598–602. doi:10.1128/aem.32.4.598-602.1976.

Immig, I. (1996). The rumen and hindgut as source of ruminant methanogenesis. *Environ. Monit. Assess.* 42, 57–72. doi:10.1007/BF00394042.

Iwamoto, M., Asanuma, N., and Hino, T. (2002). Ability of Selenomonas ruminantium, Veillonella parvula, and Wolinella succinogenes to reduce nitrate and nitrite with special reference to the suppression of ruminal methanogenesis. *Anaerobe* 8, 209–215. doi:10.1006/anae.2002.0428.

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., et al. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front. Microbiol.* 7, 1–17. doi:10.3389/fmicb.2016.00459.

Kenney, L. J., and Anand, G. S. (2020). EnvZ/OmpR two-component signaling: An archetype system that can function noncanonically. *EcoSal Plus* 9, 1–47. doi:10.1128/ecosalplus.esp-0001-2019.

Klesius, P. H. (1988). Immunity to Ostertagia ostertagi. *Vet. Parasitol.* 27, 159–167. doi:10.1016/0304-4017(88)90071-4.

Lee, C. C., Kibblewhite, R. E., Wagschal, K., Li, R., and Orts, W. J. (2012). Isolation of α-glucuronidase enzyme from a rumen metagenomic library. *Protein J.* 31, 206–211. doi:10.1007/s10930-012-9391-z.

Li, H., and O'Sullivan, D. J. (2006). Identification of a nisI promoter within the nisABCTIP operon that may enable establishment of nisin immunity prior to induction of the operon via signal transduction. *J. Bacteriol.* 188, 8496–8503. doi:10.1128/JB.00946-06.

Li, R. W., Wu, S., Li, W., Navarro, K., Couch, R. D., Hill, D., et al. (2012). Alterations in the porcine colon microbiota induced by the gastrointestinal nematode Trichuris suis. *Infect. Immun.* 80, 2150–2157. doi:10.1128/IAI.00141-12.

Lima, J., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C.-A., Snelling, T. J., et al. (2019). Identification of rumen microbial genes involved in pathways linked to appetite, growth, and feed conversion efficiency in cattle. *Front. Genet.* 10:701. doi:10.3389/fgene.2019.00701.

Love, D. N., Jones, R. F., and Bailey, A. N. D. M. (1979). Clostridium villosum sp. nov. from subcutaneous abscesses in cats. *Int. J. Syst. Bacteriol.*, 241–244.

Lynch, S. V., Goldfarb, K. C., Wild, Y. K., Kong, W., De Lisle, R. C., and Brodie, E. L. (2013). Cystic fibrosis transmembrane conductance regulator knockout mice exhibit aberrant gastrointestinal microbiota. *Gut Microbes* 4, 41–47. doi:10.4161/gutm.22430.

Maki, L. R., and Picard, K. (1965). Normal intestinal flora of cattle fed high-roughage rations. *J. Bacteriol.* 89, 1244–1249. doi:10.1128/jb.89.5.1244-1249.1965.

Malmuthuge, N., Liang, G., and Guan, L. L. (2019). Regulation of rumen development in neonatal ruminants through microbial metagenomes and host transcriptomes. *Genome Biol.* 20, 1–16. doi:10.1186/s13059-019-1786-0.

Mao, S., Zhang, M., Liu, J., and Zhu, W. (2015). Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: Membership and potential function. *Sci. Rep.* 5, 1–14. doi:10.1038/srep16116.

Martínez-Álvaro, M., Auffret, M. D., Duthie, C.-A., Dewhurst, R. J., Cleveland, M., Watson, M., et al. (2021). Bovine host genome acts on specific metabolism, communication and genetic processes of rumen microbes host-genomically linked to methane emissions. *Res. Sq. - Prepr.*, 0–37.

Martínez-Álvaro, M., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., et al. (2020). Identification of complex rumen microbiome interaction within diverse functional niches as mechanisms affecting the variation of methane emissions in bovine. *Front. Microbiol.* 11:659. doi:10.3389/fmicb.2020.00659.

Mathis, A., Weber, R., and Deplazes, P. (2005). Zoonotic potential of the microsporidia. *Clin. Microbiol. Rev.* 18, 423–445. doi:10.1128/CMR.18.3.423-445.2005.

McKellar, Q., Duncan, J. L., Armour, J., and McWilliam, P. (1986). Response to transplanted adult Ostertagia ostertagi in calves. *Res. Vet. Sci.* 40, 367–371. doi:10.1016/s0034-5288(18)30552-6.

Meyvis, Y., Geldhof, P., Gevaert, K., Timmerman, E., Vercruysse, J., and Claerebout, E. (2007). Vaccination against Ostertagia ostertagi with subfractions of the protective ES-thiol fraction. *Vet. Parasitol.* 149, 239–245. doi:10.1016/j.vetpar.2007.08.014.

Min, B. R., Gurung, N., Shange, R., and Solaiman, S. (2019). Potential role of rumen microbiota in altering average daily gain and feed efficiency in meat goats fed simple and mixed pastures using bacterial tag-encoded FLX amplicon pyrosequencing. *J. Anim. Sci.* 97, 3523–3534. doi:10.1093/jas/skz193.

Mitsumori, M., Ajisaka, N., Tajima, K., Kajikawa, H., and Kurihara, M. (2002). Detection of Proteobacteria from the rumen by PCR using methanotroph-specific primers. *Lett. Appl. Microbiol.* 35, 251–255. doi:10.1046/j.1472-765X.2002.01172.x.

Montesinos, M. L., Herrero, A., and Flores, E. (1997). Amino acid transport in taxonomically diverse cyanobacteria and identification of two genes encoding elements of a neutral amino acid permease putatively involved in recapture of leaked hydrophobic amino acids. *J. Bacteriol.* 179, 853–862. doi:10.1128/jb.179.3.853-862.1997.

Myer, P. R., Wells, J. E., Smith, T. P. L., Kuehn, L. A., and Freetly, H. C. (2015). Cecum microbial communities from steers differing in feed efficiency. *J. Anim. Sci.* 93, 5327–5340. doi:10.2527/jas.2015-9415.

Naas, A. E., Mackenzie, A. K., Mravec, J., Schückel, J., Willats, W. G. T., Eijsink, V. G. H., et al. (2014). Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio* 5, 1–6. doi:10.1128/mBio.01401-14.

Nuñez, M. (2014). Micrococcus. *Encycl. Food Microbiol. Second Ed.* 2, 627–633. doi:10.1016/B978-0-12-384730-0.00206-8.

O'Hara, E. (2019). Investigating early life microbial and host transcriptomic dynamics in the bovine gastrointestinal tract.

O'Hara, E., Neves, A. L. A., Song, Y., and Guan, L. L. (2020). The role of the gut microbiome in cattle production and health: Driver or passenger? *Annu. Rev. Anim. Biosci.* 8, 199–220. doi:10.1146/annurev-animal-021419-083952.

Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2019). Package 'vegan.'

Olaogun, S. C., and Lasisi, O. T. (2015). Bovine helminthosis: Blood glucose levels and age influence on susceptibility in some Nigerian breeds of cattle. *J. Vet. Adv.* 5, 1029. doi:10.5455/jva.20150613041006.

Ottman, N., Smidt, H., de Vos, W. M., and Belzer, C. (2012). The function of our microbiota: who is out there and what do they do? *Front. Cell. Infect. Microbiol.* 2, 104. doi:10.3389/fcimb.2012.00104.

Palarea-Albaladejo, J., and Martín-Fernández, J. A. (2020). ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach.

Patt, T. E., Cole, G. C., and Hanson, R. S. (1976). Methylobacterium, a new genus of facultatively methylotrophic bacteria. *Int. J. Syst. Bacteriol.* 26, 226–229. doi:10.1099/00207713-26-2-226.

Pernestig, A. K., Georgellis, D., Romeo, T., Suzuki, K., Tomenius, H., Normark, S., et al. (2003). The Escherichia coli BarA-UvrY two-component system is needed for efficient switching between glycolytic and gluconeogenic carbon sources. *J. Bacteriol.* 185, 843–853. doi:10.1128/JB.185.3.843-853.2003.

Petri, R. M., Schwaiger, T., Penner, G. B., Beauchemin, K. A., Forster, R. J., McKinnon, J. J., et al. (2013). Changes in the rumen epimural bacterial diversity of beef cattle as affected by diet and induced ruminal acidosis. *Appl. Environ. Microbiol.* 79, 3744–3755. doi:10.1128/AEM.03983-12.

Popova, M., McGovern, E., McCabe, M. S., Martin, C., Doreau, M., Arbre, M., et al. (2017). The structural and functional capacity of ruminal and cecal microbiota in growing cattle was unaffected by dietary supplementation of linseed oil and nitrate. *Front. Microbiol.* 8, 1–13. doi:10.3389/fmicb.2017.00937.

R Core Team (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria Available at: https://www.r-project.org/.

Rausch, S., Held, J., Fischer, A., Heimesaat, M. M., Kühl, A. A., Bereswill, S., et al. (2013). Small intestinal nematode infection of mice is associated with increased Enterobacterial loads alongside the intestinal tract. *PLoS One* 8, 1–13. doi:10.1371/journal.pone.0074026.

Reeves, P. (1995). Role of O-antigen variation in the immune response. *Trends Microbiol.* 3, 381–386. doi:10.1016/S0966-842X(00)88983-0.

Reynolds, L. A., Smith, K. A., Filbey, K. J., Harcus, Y., and James, P. (2016). Commensal-pathogen interactions in the intestinal tract: Lactobacilli promote infection with, and are promoted by, helminth parasites. *Gut Microbes* 5, 522–532. doi:10.4161/gmic.32155.Commensal-pathogen.

Roehe, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., McKain, N., Ross, D. W., et al. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* 12, e1005846. doi:10.1371/journal.pgen.1005846.

Samuel, G., and Reeves, P. (2003). Biosynthesis of O-antigens: Genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr. Res.* 338, 2503–2519. doi:10.1016/j.carres.2003.07.009.

Schieffer, K. M., Sabey, K., Wright, J. R., Toole, D. R., Drucker, R., Tokarev, V., et al. (2017). The microbial ecosystem distinguishes chronically diseased tissue from adjacent tissue in the sigmoid colon of chronic, recurrent diverticulitis patients. *Sci. Rep.* 7, 1–10. doi:10.1038/s41598-017-06787-8.

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

Settries, O. P., and Henriksen, S. A. (1989). Pneumocystis carinii in large domestic animals in Denmark. A preliminary report. *Acta Vet. Scand.* 30, 437–440. doi:10.1186/BF03548020.

Siciliano-Jones, J., and Murphy, M. R. (1989). Production of volatile fatty acids in the rumen and cecum-colon of steers as affected by forage: Concentrate and forage physical form. *J. Dairy Sci.* 72, 485–492. doi:10.3168/jds.S0022-0302(89)79130-X.

Stergiadis, S., Cabeza-Luna, I., Mora-Ortiz, M., Stewart, R. D., Dewhurst, R. J., Humphries, D. J., et al. (2021). Unravelling the role of rumen microbial communities, genes, and activities on milk fatty acid profile using a combination of omics approaches. *Front. Microbiol.* 11, 1–15. doi:10.3389/fmicb.2020.590441.

Stewart, R. D., Auffret, M. D., Warr, A., Wiser, A. H., Press, M. O., Langford, K. W., et al. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* 9, 870. doi:10.1038/s41467-018-03317-6.

Strom, E., and Øskov, E. R. (1984). The nutritive value of rumen micro-organisms in ruminants. *Br. J. Nutr.* 52, 613–620. doi:10.1079/bjn19840128.

Summerbell, R. C. (2004). “Fungi associated with vertebrates,” in *Biodiversity of fungi: Inventory and monitoring methods* (Elsevier Inc.), 451–465. doi:10.1016/B978-012509551-8/50023-4.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., et al. (2009). Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genet.* 5. doi:10.1371/journal.pgen.1000344.

Valladares, A., Montesinos, M. L., Herrero, A., and Flores, E. (2002). An ABC-type, high-affinity urea permease identified in cyanobacteria. *Mol. Microbiol.* 43, 703–715. doi:10.1046/j.1365-2958.2002.02778.x.

van Zijderveld, S. M., Gerrits, W. J. J., Apajalahti, J. A., Newbold, J. R., Dijkstra, J., Leng, R. A., et al. (2010). Nitrate and sulfate: Effective alternative hydrogen sinks for mitigation of ruminal methane production in sheep. *J. Dairy Sci.* 93, 5856–5866. doi:10.3168/jds.2010-3281.

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C.-A., Hyslop, J. J., Ross, D. W., et al. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 16, 839. doi:10.1186/s12864-015-2032-0.

Weyrich, L. S., Feaga, H. A., Park, J., Muse, S. J., Safi, C. Y., Rolin, O. Y., et al. (2014). Resident microbiota affect bordetella pertussis infectious dose and host specificity. *J. Infect. Dis.* 209, 913–921. doi:10.1093/infdis/jit597.

Willey, J. M., Sandman, K. M., and Wood, D. H. (2020a). “Gram-Positive bacteria,” in *Prescott’s Microbiology* (New York), 537–558.

Willey, J. M., Sandman, K. M., and Wood, D. H. (2020b). “Proteobacteria,” in *Prescott’s Microbiology* (New York), 504–536.

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

Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36, 808–812. doi:10.2144/04365st04.

Chapter 7 General discussion

7.1 Introduction

Bovine production systems are presently under more pressure than ever before because they must tackle current goals associated with worldwide food security, and pollution, as well as animal health and welfare. The study of the taxonomic composition and functional potential of gastrointestinal microbiomes in association with economically compelling traits of the animal host, such as feed conversion efficiency, appetite, growth rate, and resilience to nematode parasitism, is essential for the progress of improved production strategies, whether it be related to nutrition, breeding programmes, or development of vaccines. This thesis focuses on several aspects of the gastrointestinal microbiome, including the use of different bioinformatics pipelines to resolve the microbial communities composition of samples based on the 16S rRNA gene derived from swine gastrointestinal samples, the impact of bovine rumen microbial genes profiles on performance traits including feed conversion efficiency, appetite, and growth, the temporal stability of the rumen microbiome during the finishing phase of beef cattle and its impact on the prediction of performance and methane emissions traits, and the influence of an abomasal parasitic nematode on the rumen and caecum microbiomes of dairy cattle. In this general discussion, I will firstly describe some of the most important results of each of our projects, and then discuss them in a wider context, elaborating on their implications, and highlighting their strengths and limitations.

7.2 Application of different bioinformatics pipelines leads to different microbiota profiles

Studies focused on the composition of the microbiome rely heavily on the software used. Bioinformatics tools are crucial for microbiome studies, not only due to the huge amount of data available to analyse, but also due to its complexity. The accurate taxonomic characterization of microbiota in samples can be achieved through different strategies. At present, the most used

techniques involve the processing of either 16S rRNA gene amplicons, or of contigs obtained through next generation whole genome shotgun sequencing.

7.2.1 16S rRNA gene-based strategies

The use of the 16S rRNA gene in the study of microbial communities in gastrointestinal samples is widely recognized as a sound strategy. This gene is present in the genome of most prokaryotes, but it varies enough between groups to be used in their distinction. The 16S rRNA gene is a housekeeping gene, and since its function has not changed over time, the changes (i.e., mutations) it has suffered are believed to be associated with evolution, therefore mirroring the phylogeny between organisms. Techniques based on the 16S rRNA gene are attractive, in comparison to standard microbiology identification techniques, mostly due to the possibility of identifying large numbers of microorganisms in the sample that would be costly and sometimes impossible to grow on culture plates following standard laboratory procedures, and also because they produce more accurate and objective results (Janda and Abbott, 2007; Petti, 2007). Additionally, in comparison to next generation sequencing techniques, they are fast and substantially less costly.

In chapter 2, we summarize the characteristics of two bioinformatics pipelines widely used to resolve taxonomic compositions based on the 16S rRNA gene, the MetaGenome Rapid Annotation using Subsystem Technology, and the Quantitative Insights Into Microbial Ecology 2 (MG-RAST, and QIIME2, respectively). Additionally, we compared the taxonomic compositions obtained from each of these pipelines and discussed them considering the algorithms employed by each pipeline.

The main differences between the algorithms underlying MG-RAST and QIIME2 are related with the process of grouping 16S rRNA gene amplicons based on their similarity. Whereas MG-RAST groups amplicon reads based on a 97% similarity threshold, creating operational taxonomic units (OTUs), QIIME2 employs an algorithm that groups sequences into amplicon sequence variants (ASVs) based on an error model that accounts for the increasing error

rates along the length of the sequence, the number of copies of unique sequences, and the similarity between unique sequences.

Processing the 188 samples collected from the swine gastrointestinal tract in MG-RAST and QIIME2 led to different microbiota compositions. At domain level, MG-RAST erroneously classified a higher number of amplicons as Eukaryota and Viruses than QIIME2, and a higher percentage of hits was left unclassified. At phylum level, MG-RAST and QIIME2 identified 15 taxa in common (accounting for 100% and 98.8% of the total hits, respectively), whereas 7 extra taxa were identified exclusively by QIIME2, but accounted only for a small portion of the total microbiota. At family level, *Prevotellaceae* and *Ruminococcaceae* were the 1st and 2nd most abundant taxa in both pipelines, however, the 3rd, 4th, and 5th most abundant families were, respectively, *Veillonellaceae*, *Clostridiaceae*, and *Eubacteriaceae*, in MG-RAST, and *Lachnospiraceae*, *Muribaculaceae*, and *Veillonellaceae*, in QIIME2.

The different microbiota compositions obtained from using MG-RAST and QIIME2 impacted the downstream statistical analyses. To evaluate this impact, we used partial least squares (PLS) models to discriminate between microbiota compositions derived from samples collected from the caecum, colon, and faeces. Several cleaning and filtering procedures were also tested within these analyses. The results revealed that filtering out microbial genera with low average relative abundance led to QIIME2-derived microbiota profiles being more accurate for the discrimination between sample types, whereas when including low abundance taxa in the matrix of explanatory variables, MG-RAST resulted in higher accuracy. It follows that, if the low abundance microbial taxa are real, MG-RAST would lead to a more accurate characterization of the microbiota profiles, whereas, if these are false positives, QIIME2 would be preferable. ASV-based pipelines, such as QIIME2, may lead to an artificial increase in false positives, and consequently to an artificially increase in samples' diversity, and therefore the application of a minimum

relative abundance threshold may be beneficial for the analyses of the obtained datasets.

This study showed that there are differences on microbiome profiles obtained using different pipelines and elucidated some of the features in each pipeline that are on the basis of these differences, providing useful guidelines for potential users.

7.2.2 Next-generation sequencing-based techniques

Although there are multiple advantages in the use of 16S rRNA gene for the characterization of microbiota profiles, the application of whole genome shotgun sequencing (WGS) methods is the natural evolution of the field. One of the most important features of WGS is the ability of deriving the composition of the microbiome at the microbial genes level, facilitating the analyses of the functional potential of the microbiome in a direct manner, rather than by inference-based methods possible with the 16S rRNA gene. Additionally, WGS methods improve on the accuracy of the microbiota characterization, making it possible to resolve microbiome samples even at the species level if genomics of the microbes in the samples are well characterized in databases. Considering these advantages, the studies presented in chapters 3, 4, 5, and 6 within this thesis were based on microbiome samples characterized at the microbial genera and microbial genes levels, using WGS data.

7.3 Bovine performance traits are closely associated with rumen microbial genes

The second main aim of this thesis was to understand the association between the functional profiles of the rumen microbiome with performance traits in beef cattle, including appetite, growth rate, and feed conversion efficiency. In chapter 3, I present our published manuscript, in which we identified sets of 20, 14, 17, and 18 microbial gene biomarkers which explained 63, 65, 66, and 73% of the variation in FCR, ADG, RFI, and DFI. We obtained these results by using PLS to identify the microbial genes whose relative abundances were closely associated with each performance trait. The results showed that

specific microbial gene biomarkers were mostly exclusive associated with each trait, reflecting the different microbial functional networks associated with different host traits. However, some microbial genes were commonly identified as biomarkers for highly correlated traits, such as FCR with ADG and RFI with DFI.

Microbial genes associated with FCR were involved in carbohydrates metabolism and transport, such as xylan degradation, and pentose and glucuronate interconversions. Microbial genes important for the prediction of ADG were involved in peptidoglycan turnover, in association with cAMP resistance, typical of pathogenic organisms, which had relative higher abundance in animals with lower ADG. In relation to RFI biomarkers, we identified several microbial genes associated with chemotaxis, and motility in microorganisms. The microbial genes more important for the prediction of DFI were associated with lipopolysaccharide (LPS), a major virulence factor in Gram-negative bacteria.

Microbial genes involved in bacterial defence mechanisms were relatively more abundant in less efficient animals, which could be indicative of pathogens present in the rumen, representing a potential energy sink for the host. In highly efficient animals, higher relative abundances of microbial genes associated with cellulose and hemicellulose degradation, amino acids metabolism, and vitamin B12 biosynthesis could be indicative of improved fermentation processes.

7.4 Temporal stability of the rumen microbiome

The association of the microbiome with host traits in the study presented in chapter 3 (microbial genes in association with host performance traits) was based on the characterization of the rumen microbiome derived from WGS of samples collected at slaughter. However, whether slaughter samples are representative of the rumen microbiome during the bovine finishing phase was still unclear. Therefore, in chapter 4, we present a comparison of microbiome profiles of samples collected from the rumen before including a nitrate- or oil-

based diet additive, at the start, mid, and end of a 56-day performance testing period, after the animals left the chamber in which they were individually measured for methane emissions, and at slaughter.

This study showed that the rumen microbiome, both at taxonomic and microbial genes level, is highly stable throughout the finishing phase of beef cattle. For example, the regression of ALR-transformed abundances of microbial genera and microbial genes on sampling timepoint, including animal ID as random effect revealed that 99% of the microbiome traits showed no significant differences throughout time of sampling. Additionally, the Pearson correlations between vectorized matrices of microbial genera and microbial genes in samples collected at different timepoints were highly correlated, with the lowest being between pre-additive and end test (70.58% and 81.36 for microbial genera and microbial genes, respectively).

We also investigated the temporal stability of the associations of the rumen microbiome with performance traits, including FCR, ADG, DFI, RFI, CH_4 yield and CH_4 production. These analyses showed that the microbiome at microbial genera or microbial genes level is highly associated with the traits throughout the finishing phase. Additionally, there was a substantial agreement of the microbial genera and genes that were identified as important for explaining the variation of the performance traits based on slaughter data, and those based on previous sampling timepoints. This result is consistent with our shown previous finding that sampling timepoint had no effect on the microbiome datasets generated from samples collected at different times. These results suggested that microbiome profiles derived from slaughter samples are representative of the microbiome throughout the whole phase, and can therefore be used in studies that characterize the rumen microbiome, or in the identification of biomarkers of host traits, as we had previously done in the published paper “Identification of Rumen Microbial Genes Involved in Pathways Linked to Appetite, Growth, and Feed Conversion Efficiency in Cattle”, presented in chapter 3. In this context, in chapter 4, we specifically verified whether the microbial genera and microbial genes previously identified

as biomarkers for host traits (from the published paper in chapter 3, and from other published papers), would still have been identified if the analyses had been based on data derived from earlier sampling timepoints. The results confirmed that the associations between microbial genera and microbial genes identified as biomarkers are highly stable throughout the finishing phase of beef cattle.

Furthermore, the stability of associations between host-genomically influenced microbial genes (i.e., microbial genes with moderate to high heritability) and the estimated breeding values (EBVs) of performance traits was also tested. The results showed that the analyses based microbiome profiles obtained in samples collected at earlier timepoints mostly agreed with those based on slaughter-derived data.

7.5 The rumen microbiome in association with bovine health

The bovine gastrointestinal microbiome has been increasingly recognized as central for improvements in productivity (e.g., feed conversion efficiency and appetite), and environmental impact (e.g., methane emissions). The relationship between the gastrointestinal microbiome and the animal host has been reported to also influence the host health. For example, Nagaraja & Titgemeyer (2007) elucidated the impact of diet composition on the rumen microbiota, and how excess fermentable carbohydrates can lead to ruminal acidosis, and Gomez et al. (2019) discussed how the immune system development in calves is dependent on the proper establishment of healthy microbiota communities in the gastrointestinal tract.

The fourth main aim of this thesis was to understand whether parasitism by the abomasal nematode *Ostertagia ostertagi* impacted the rumen and caecum microbiome in dairy calves, and if so, which are the microbial genera and microbial genes mostly affected by the parasitism. We did this by comparing the microbiome profiles of dairy calves subjected to an infection challenge that consisted of oral administration of 1000 L3 larvae per day for 25 days, with

those of animals in a negative control group (unchallenged). Additionally, we investigated the impact of vaccinating the animals with a native vaccine against *O. ostertagi*, by comparing the rumen and caecum microbiome profiles of vaccinated animals (vaccinated and challenged with infection), with those of animals in a negative control group (unvaccinated and unchallenged), as well as with those of infected animals (positive control, unvaccinated, challenged).

The influence of the parasite and the vaccine was investigated on the rumen and caecum microbiome profiles separately, due to the different functions of these organs in the digestive processes of the ruminant. Specifically, the rumen is responsible for the pre-gastric digestion of the feed, whereas in the caecum feed is digested that has escaped the rumen undigested. Because of the location of these organs in the gastrointestinal system, and of the type of substrates they receive, their microbiota is substantially different, resulting in different fermentation capabilities. These differences are explored in chapter 6.

7.6 Differences between the rumen and caecum microbiome profiles

The data used in chapters 5 and 6 offered us the unique opportunity of comparing microbiome profiles of rumen and caecum. The results are presented in chapter 6. One of the most striking differences observed was the increased relative abundance of Bacteroidetes and Proteobacteria in the rumen, and the increased relative abundance of Actinobacteria and Firmicutes in the caecum, reflecting the specific fermentation profiles of these organs, since the diet entering the rumen is richer in easily fermentable carbohydrates than the diet received by the caecum.

Although the rumen microbiome is associated with the majority of methane emissions produced by the animal, the caecum microbiome has also been shown to contribute with between 6 and 14% of the daily methane production (Immig, 1996). Interestingly, although the Archaea:Bacteria ratio in the rumen has been shown to be an accurate biomarker for methane production (Roehe

et al., 2016), we observed no significant differences in relative abundances of Euryarchaeota when comparing caecum and rumen microbial profiles. This suggests that the lower contribution of the caecum microbiome to the methane production may be associated with different fermentation profiles than those found in the rumen microbiome, since easily fermentable substrates associated with high hydrogen release, closely associated with methanogenesis in the rumen, are virtually used up in the rumen, and never reach the caecum. In agreement, hydrogen-producing fibrolytic bacteria, including *Prevotella*, *Succiniclasticum*, and *Fibrobacter* were the dominant microbial genera in the rumen, as well as hydrogen-utilizer *Methanobrevibacter* (Shinkai et al., 2010; Mi et al., 2018).

7.7 The rumen and caecum microbiome profiles are affected by the presence of the abomasal nematode *Ostertagia ostertagi*

The comparison of rumen microbiome profiles of infected with uninfected animals revealed depleted abundances of microbial genera associated with beneficial fermentation processes. For example, *Butyrivibrio* and *Pseudobutyrivibrio*, depleted in both infected groups (infected showing high or low ncumulative faecal egg count, CHE, and CLE, respectively), are important utilizers of xylans and pectins and other carbohydrates in the production of VFAs, supplying energy for the ruminant. In parallel, we observed enrichment of opportunistic pathogens in microbiome profiles of infected animals, including e.g., *Streptomyces*, *Tsukamurella*, *Pseudopropionibacterium* and *Segniliparus*.

The investigation of the differences regarding the functional profiles of the rumen microbiome in these animals showed that infected animals were depleted of microbial genes associated with amino acids biosynthesis and metabolism (e.g., involved in the histidine metabolism, and arginine biosynthesis), which could be associated with the hindered growth rate reported by previous authors (Fox et al., 1989). Additionally, we explored how the microbial gene biomarkers identified in the published paper of chapter 3

(Lima et al., 2019) were affected by *O. ostertagi* parasitism. The results showed that microbial gene biomarkers previously associated with decreased feed conversion efficiency, and decreased growth rates were enriched in infected animals, whereas microbial gene biomarkers previously associated with increased feed conversion efficiency and increased appetite were here depleted in infected animals.

Additionally, infected animals had their rumen microbiomes enriched with microbial genes associated with bacterial cell wall structures, environmental sensing processes, and bacterial defence mechanisms.

In chapter 6, we explored the impact of *O. ostertagi* parasitism on the caecum microbiome of dairy calves. The results showed increased abundance of opportunistic pathogens such as *Babesia* and *Toxoplasma* in the caecum of infected, in comparison to uninfected animals. However, other opportunistic pathogens were depleted in the caecum of infected animals, including several Gammaproteobacteria, such as *Acitenobacter* and *Legionella* and Bacilli, e.g., *Ehrlichia*. These are unexpected results, since we had observed enrichment of pathogenic organisms in the rumen of infected animals (chapter 5). In line with the expectation, favourable microbial genera such as acid lactic producers *Lactobacillus* and *Tetragenococcus* were depleted in the caecum of infected in comparison to uninfected animals.

7.8 Influence of a native vaccine against *O. ostertagi* parasitism on the rumen and caecum microbiome profiles

Interestingly, the vaccine in both the rumen and the caecum microbiomes led to increased abundances of fungi belonging to phyla Ascomycota and Basidiomycota, most of them with pathogenic potential, in comparison to uninfected animals. Additionally, the comparison of microbiome profiles at genes level suggested that both in the rumen and in the caecum, the vaccinated animals had increased abundance of microbial genes associated with environmental sensing, cell wall structure and flagellar assembly.

The abundance of opportunistic pathogens, such as *Bartonella* and *Bosea* was depleted in the rumen microbiome of vaccinated, in comparison to infected animals, suggesting that the vaccination prevented some of the dysbiosis associated with the parasitism. Additionally, the caecum of vaccinated animals was enriched in opportunistic pathogen *Bordetella*, and in members of the *Enterobacteriaceae* family (e.g., *Enterobacter*), previously shown to be associated with nematode infections in mice (Reynolds et al., 2016).

In the caecum of vaccinated animals, we also observed enrichment of methylothrophic *Methyloceanibacter*, previously shown to be negatively correlated with methane emissions in beef cattle (Auffret et al., 2018), and sulfur-reducing acetate producer *Desulfovibrio* (Howard and Hungate, 1976; Willey et al., 2020). Since sulfate reduction is an effective hydrogen sync, the enrichment of sulfate reducing organisms together with methylotrophs such as *Methyloceanibacter* may influence methane emissions in vaccinated animals, however, methane emissions were not measured within this study.

7.9 Challenges in microbiome-centric studies

Studies focused on the characterization of microbiome datasets, and their association with host traits, face several challenges, which I will briefly summarize and discuss here, as they were in continuous development for microbiome data during this PhD project and beyond.

7.9.1 The “large p small n” challenge

The first main challenge when working with microbiome or other -omics datasets stems from their inherent complexity. Microbiome datasets often include hundreds or even thousands of variables (e.g., microbial genera or microbial genes), and this high dimensionality, together with very complex biological interrelationships between the variables, is one of the main challenges in the interpretation of results. When the number of observations (n) is lower than the number of predictor variables (p), ordinary least squares cannot be applied, because n has to be greater than the rank of the covariance

matrix (of dimensions p × p), otherwise the covariance matrix is singular. However, when n is lower than p, PLS models can be applied.

PLS models are constructed based on the singular value decomposition (SVD) of a matrix containing the continuous response variables (i.e., Y), and of a matrix containing the continuous predictor variables (i.e., X). The SVD of the Y matrix creates a lower rank matrix, meaning that the information contained on the original Y is described by fewer components, in a strategy like that of a principal component analysis (PCA), following the equation $Y = UQ' + F$, in which U refers to the scores, Q refers to the loadings, and F refers to the residuals of the SVD. The matrices U and Q are orthogonal, i.e., they are statistically independent of each other. The loadings Q can be understood as the weights of each original variable in the construction of the latent components, whereas the scores U are the original data rotated to the new coordinates system. It follows that, after SVD, Y can be estimated based on U. Similarly, the SVD of the X matrix follows the equation $X = TP' + E$, in which T are the scores, P are the loadings and E are the residuals. The main difference between PCA and PLS is that whereas PCA will work on a sole matrix of variables, creating latent components (i.e., principal components) that will maximize the variance of the matrix, PLS will work with two matrices, and will create latent variables that maximize the covariance between them. Therefore, in the PLS models, the dimension reduction (i.e., the SVD of the predictor and response matrices) occurs simultaneously with a regression, in which scores U and T and loadings Q and P are developed so that U and T have maximum covariance. It follows that, since Y can be estimated from U, and U and T are developed as to have maximum covariance, Y can be estimated from X (Boulesteix and Strimmer, 2007).

One of the most interesting features of PLS models is the variable importance in projection (VIP) scores. These are defined for each variable in the predictor matrix and reflect the importance of each predictor variable for the prediction of the dependent variable. VIPs are calculated based on the weight of each variable on each PLS latent component, weighted by the percentage of Y

variation explained by each PLS latent component. Since the sum of squared VIP scores for all variables equals the number of variables, the VIP scores vary within a fixed range, and have average 1 (Cocchi et al., 2018). Therefore, variables with VIP ≥ 1 (variables with VIP higher than the average squared VIP) are often identified as important for the prediction of the response variable (Sun et al., 2016; Zeng et al., 2019; Wang et al., 2021). Some other authors have applied a threshold of VIP ≥ 0.8, to be less strict (Wallace et al., 2015; Roehe et al., 2016; Auffret et al., 2018; Martínez-Álvaro et al., 2020). In chapters 3, 4, 5 and 6, the VIP threshold applied was 1.

PLS models can also be used when the response variable is categorical (i.e., partial least squares discriminant analysis, PLS-DA). The use of PLS-DA to discriminate groups of observations has been previously criticized, because this classification tool is based on a PLS model in which the dependent variable is chosen to represent the group membership, in an analyses with a large number of variables and a low number of observations. Westerhuis et al. (2008) stated that “PLS-DA is eager to please and thus its results should be handled with great care”, particularly because the model is forced to infer class separation. Additionally, the authors criticize the use of AUROC and other forms of cross-validation (e.g., number of misclassifications) due to the inexistence of a gold standard value against which we could compare these results. Therefore, besides using the AUROC and the prediction error rate in the explorative analyses of chapters 5 and 6, we also used PCA because PCA models do not include any information regarding the “expected grouping” of the observations, and therefore, if a cluster patter becomes apparent, it is solely due to the predictor variables matrix.

PLS and PLS-DA were performed in the studies of chapter 3, 5 and 6 in an iterative manner, in which the least important variables (VIP < 1) of the first model were removed from the second model, and so forth. This procedure was applied because, although factors that alter the structure of the microbiome will most likely affect all microbial communities (and therefore microbial genes), due to the large number of variables available in each observation

some of these variables will change more radically than others, and our goal was to identify the microbiome features more strongly associated with performance traits (chapter 3), and more strongly influenced by *O. ostertagi* parasitism and by the vaccine (chapters 5 and 6).

7.9.2 The compositional nature of microbiome datasets

Microbiome datasets are inherently compositional. As evidenced in Gloor et al. (2017), by using a curious comparison between datasets resulting from high-throughput sequencing (HTS) with that of an ecological study scenario involving counts of tigers and ladybugs. In HTS experiments, there is a maximum number of reads obtained from each sample, which is defined by the capacity of the sequencing instrument used; when counting tigers and ladybugs in a given location, there is no such maximum number of counts. Therefore, HTS techniques result in non-independence of counts between different groups, whereas the counts of tigers will be statistically independent from that of ladybugs, i.e., the counts of a given variable in an HTS-derived dataset may be predicted from the counts of the other variables, whereas the number of tigers cannot be predicted from the number of ladybugs. Furthermore, there is no apparent interest in the total number of counts in a sample, but rather in the relations between these variables.

One of the most important principles in compositional data analysis is called subcompositional coherence. This means that results obtained from statistics using a subcomposition (i.e., a subset of the variables) should not differ from those obtained from the original composition (i.e., considering all variables). The normalization of counts by the total number of counts in a sample (i.e., relative abundance) is often applied in microbiome-centric studies and was applied in chapters 2 and 3. This process, called closure, is often performed to avoid dilution effects in the samples. The normalization by the total number of counts leads therefore to a matrix in which any part can be predicted from the other parts (because they all sum 1). However, in the matrix of relative abundances, the principle of subcompositional coherence is not respected; considering for example a composition of 4 parts (variables A to D), and a

subcomposition including variables A and B only, the correlation between variables A and B based on the complete composition will differ from the correlation calculated between A and B based on the subcomposition.

To overcome the issues of using matrices of relative abundances directly in statistical analyses, there are some transformations options, of which we will discuss the additive logratio (ALR) and the centred logratio (CLR). The ALR is calculated as the logarithm of the ratio between each variable (i.e., abundance of each genus or microbial gene) and a selected denominator (i.e., the abundance of a selected microbial genus or gene with specific criteria). The CLR is the logarithm of the ratio of the counts of each variable by the geometric mean of all variables in the sample.

As explained in Greenacre et al. (2021), although ALRs are not isometric, i.e., they do not preserve the original distance between observations, this small loss of isometry is compensated by a simpler and clearer interpretation of the logratio variables, in comparison to other, isometric, method. When applying the ALR transformation, the denominator must be selected as to maximize the Procrustes correlation between the (new) ALR-transformed matrix and the original CLR-transformed matrix, as to preserve the total logratio variance. Of the variables that satisfy this first condition, the ones with lowest variance should be the best candidates to become denominator, for two main reasons; first, using a non-variant denominator means that all values are just shifted by an almost constant value (since the log(A/B) = log(A) – log(B)), and therefore the ALR-transformed values can be easily interpreted independently of the denominator, i.e., if the value of an ALR-transformed variable increases from one observation to the next, this is directly interpreted as an increase in the numerator of the ratio, and never as a decrease in the denominator, since the denominator is (almost) the same; second, if a variable shows no variation between different observations, it is unlikely that it is importantly associated with whichever trait we are exploring, and therefore by selecting it as denominator (and not using it as a predictor variable) we are not losing information. A further criterion for the selection of the denominator is the

potential interest in this variable as a predictor of the response variable, considering the biological context in which these transformations are used. We have applied the ALR transformation in chapter 4.

The CLR transformation depends on the geometric mean of all variables in a sample, making the CLR variables linearly dependent (each part can be predicted from all other parts of the full composition). However, in comparison to using relative abundances, this method is still preferable, since it brings compositional data from a simplex universe (i.e., bounded by the total sum of parts) into the universe of real numbers (i.e., the Euclidean space), in which classical statistics can be applied (Greenacre, 2018). The CLR method is extremely useful for computational purposes, and it was used here in chapters 5 and 6.

7.10 The wider context of these results, and their implications

The results of this thesis, discussed in the previous sections of this chapter, give insight into several perspectives of the associations between the gastrointestinal microbiome and the animal host, with impact on performance traits, and health.

We have identified microbial gene biomarkers for feed conversion efficiency, appetite, and growth, in beef cattle (chapter 3). One of the most important possible applications of these results is the use of these biomarkers for breeding purposes and to investigate the impact of dietary interventions without the costly measurement of these traits, in particular feed intake. Furthermore, our study on the temporal stability of the rumen microbiome (chapter 4) showed that the rumen microbiome and its association with host performance traits, including feed conversion ratio, growth rate, and methane production, is highly stable during the finishing phase of beef cattle. Considering the results of both these chapters, the animals with highest feed conversion efficiency and lowest methane emissions could be identified based on their microbiome profiles even based on the start of the finishing period.

One of the main caveats of our results is that they were mainly obtained based on phenotypic level and should be further analysed on the genetic level. This was not possible on the used data because it was not sufficient for a genetic analysis. Therefore, this study was aligned with a project using a substantially larger dataset to estimate genetic parameters and EBVs (Martínez-Álvaro et al., 2021) including the animals with longitudinal microbiome information. The analyses of this longitudinal microbiome data in chapter 4 revealed strong associations between abundances of microbial genes with moderate to high heritability and the EBVs of performance traits during the finishing phase of cattle. This result is of extreme importance, due to its potential implications in the selection of animals for breeding. Since heritable microbial genes are highly stable throughout time, the selection of animals for breeding can be performed based on microbiome profiles derived from samples taken as early as the start of the finishing phase. Furthermore, since heritable microbial genes were shown here to have a stable association with the EBVs of the performance traits, it is possible to select high-performance animals for breeding based on the microbial genes, and therefore use them as predictors of breeding values, or as complementary information in the prediction of EBVs, which has been shown to increase the accuracy of the estimates Martínez-Álvaro et al. (2021).

The knowledge of the function of the microbial genes highly associated with increased host animal performance enlightens on the microbial functional networks underlying these economically important performance traits, which potentially informs on the requirements for the development of potential probiotics, or individualized targeted nutritional interventions, that can lead to further improvement of animals' performance as well as health and welfare.

The gastrointestinal microbiome is not only closely associated with the host animals' performance, but also with their health, as shown in chapters 5 and 6. *Ostertagia ostertagi* is one of the most economically impactful parasitic nematodes in bovine production systems, leading to inappetence and stunted

growth of infected animals. In the most developed systems, the effects of this parasite are counteracted using antelminths.

Previous studies on how *O. ostertagi* causes disease in animals, particularly in ruminants, have clarified on some of the direct effects of the parasite within the host organism, such as the damage to the gastric glands as the young adult parasite exits them, associated with decreased hydrochloric acid production, leading to increased pH, hypergastrinemia, and hyperpepsinogenaemia. However, the effects of this abomasal parasite on the microbiome in the rumen and caecum were, to the best of our knowledge, not investigated using microbiome profiles based on whole metagenomic sequenced data. Since the nematode has a high impact on animal performance and health traits, it is of great interest whether this may be at least partly due to changes in the ruminal and caecal microbiome profiles. Therefore, we explored the impact of the parasite and of a native vaccine against the parasite, on the rumen and caecum microbiome (chapters 5 and 6). This research complements those in chapters 2 to 4, where we have shown that microbial genes in the rumen are highly associated with performance traits, including appetite, and growth. Therefore, it can be suggested that the dysbiosis in the rumen and caecum microbiome profiles observed in infected animals may also have resulted in inappetence and stunted growth of parasitised animals.

One of the reasons parasitism by *O. ostertagi* has a severe impact on animal health is that immunity to this parasite, unlike other parasitic nematodes of ruminants, takes a long time to develop, mainly affecting young animals up to 2 years old. The negative consequences of inappetence and stunted growth during the initial growth stage of animals can have repercussions into their growth rate and productivity during mature stages (Everitt and Jury, 1977; Perri et al., 2011), which means that the consequences of early-life parasitism can extend throughout the whole life of the animals' productive life span.

The potential development of nematodes resistant to anthelmintics makes alternative strategies more compelling. The only well-established alternative strategy for nematode control is based on grazing management, which aims at limiting the contact between the host and the parasite (Charlier et al., 2009), though vaccines are also being developed. The identification of microbial genera and functional genes potentially affected by the parasitism by *O. ostertagi* may pose an opportunity to develop a new alternative strategy or complement existing ones; since the rumen and the caecum microbiome profiles are closely associated with the host performance traits and animal health and welfare, and intervention on these microbial ecosystems may help circumvent the negative consequences of parasitism on host performance, particularly while they develop their own immunity response. For example, breeding for more resilient animals to *O. ostertagi* could be based on the obtained rumen biomarkers identified in chapter 5 to reduce the adverse effects of ostertagiasis on the animal, including host-genomically influenced microbial gene biomarkers depleted in the rumen of infected animals, such as *argO* (i.e., N-acetylglutamate synthase), involved in biosynthesis of amino acid arginine, ME2 (i.e., malate dehydrogenase (oxaloacetate-decarboxylating)), involved in carbon metabolism, and *pdp* (i.e., pyrimidine-nucleoside phosphorylase), involved in pyrimidine metabolism.

Based on all results of this thesis, it can be concluded that the rumen microbiome is highly informative for estimation of performance and methane emissions traits, as well as animal health. This can be explored, e.g., by breeding, and dietary interventions including the use of probiotics or prebiotics, as well as in the development of vaccines with minimal impact on the gastrointestinal microbiome. In particular, the profiles of microbial genes among animals were shown to be very informative, because the abundances of these genes provided information about the functional network of the microbiome associated with performance and methane emissions traits, as well as animal health. Based on this information, many potential microbial

functions associated with these traits were identified and can be verified and extended in future studies.

7.11 References

Auffret, M. D., Stewart, R., Dewhurst, R. J., Duthie, C.-A., Rooke, J. A., Wallace, R. J., et al. (2018). Identification, comparison, and validation of robust rumen microbial biomarkers for methane emissions using diverse Bos taurus breeds and basal diets. *Front. Microbiol.* 8, 1–15. doi:10.3389/fmicb.2017.02642.

Boulesteix, A. L., and Strimmer, K. (2007). Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Brief. Bioinform.* 8, 32–44. doi:10.1093/bib/bbl016.

Charlier, J., Höglund, J., von Samson-Himmelstjerna, G., Dorny, P., and Vercruysse, J. (2009). Gastrointestinal nematode infections in adult dairy cattle: Impact on production, diagnosis and control. *Vet. Parasitol.* 164, 70–79. doi:10.1016/j.vetpar.2009.04.012.

Cocchi, M., Biancolillo, A., and Marini, F. (2018). “Chemometric methods for classification and feature selection,” in *Comprehensive Analytical Chemistry* (Elsevier B.V.), 265–299. doi:10.1016/bs.coac.2018.08.006.

Everitt, G. C., and Jury, K. E. (1977). Growth of cattle in relation to nutrition in early life. *New Zeal. J. Agric. Res.* 20, 129–137. doi:10.1080/00288233.1977.10427316.

Fox, M. T., Gerrelli, D., Pitt, S. R., Jacobs, D. E., Gill, M., and Gale, D. L. (1989). Ostertagia ostertagi infection in the calf: effects of a trickle challenge on appetite, digestibility, rate of passage of digesta and liveweight gain. *Res. Vet. Sci.* 47, 294–298. doi:10.1016/s0034-5288(18)31249-9.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* 8, 1–6. doi:10.3389/fmicb.2017.02224.

Gomez, D. E., Galvão, K. N., Rodriguez-Lecompte, J. C., and Costa, M. C. (2019). The cattle microbiota and the immune system: An evolving field. *Vet. Clin. North Am. - Food Anim. Pract.* 35, 485–505. doi:10.1016/j.cvfa.2019.08.002.

Greenacre, M. (2018). *Compositional data analysis in practice.* , ed. C. & H. / C. Press.

Greenacre, M., Martínez-Álvaro, M., and Blasco, A. (2021). Compositional data analysis of microbiome and any-omics datasets : a revalidation of the additive logratio transformation. doi:doi: https://doi.org/10.1101/2021.05.15.444300.

Howard, B. H., and Hungate, R. E. (1976). Desulfovibrio of the sheep rumen. *Appl. Environ. Microbiol.* 32, 598–602. doi:10.1128/aem.32.4.598-602.1976.

Immig, I. (1996). The rumen and hindgut as source of ruminant methanogenesis. *Environ. Monit. Assess.* 42, 57–72. doi:10.1007/BF00394042.

Janda, J. M., and Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J. Clin. Microbiol.* 45, 2761–2764. doi:10.1128/JCM.01228-07.

Lima, J., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C.-A., Snelling, T. J., et al. (2019). Identification of rumen microbial genes involved in pathways linked to appetite, growth, and feed conversion efficiency in cattle. *Front. Genet.* 10:701. doi:10.3389/fgene.2019.00701.

Martínez-Álvaro, M., Auffret, M. D., Duthie, C.-A., Dewhurst, R. J., Cleveland, M., Watson, M., et al. (2021). Bovine host genome acts on specific metabolism, communication and genetic processes of rumen microbes host-genomically linked to methane emissions. *Res. Sq. - Prepr.*, 0–37.

Martínez-Álvaro, M., Auffret, M. D., Stewart, R. D., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., et al. (2020). Identification of complex rumen microbiome interaction within diverse functional niches as mechanisms affecting the variation of methane emissions in bovine. *Front. Microbiol.* 11:659. doi:10.3389/fmicb.2020.00659.

Mi, L., Yang, B., Hu, X., Luo, Y., Liu, J., Yu, Z., et al. (2018). Comparative analysis of the microbiota between sheep rumen and rabbit cecum provides new insight into their differential methane production. *Front. Microbiol.* 9, 1–14. doi:10.3389/fmicb.2018.00575.

Nagaraja, T. G., and Titgemeyer, E. C. (2007). Ruminal acidosis in beef cattle: The current microbiological and nutritional outlook. *J. Dairy Sci.* 90, E17–E38. doi:10.3168/jds.2006-478.

Perri, A. F., Mejía, M. E., Licoff, N., Lazaro, L., Miglierina, M., Ornstein, A., et al. (2011). Gastrointestinal parasites presence during the peripartum decreases total milk production in grazing dairy Holstein cows. *Vet. Parasitol.* 178, 311–318. doi:10.1016/j.vetpar.2010.12.045.

Petti, C. A. (2007). Detection and identification of microorganisms by gene amplification and sequencing. *Clin. Infect. Dis.* 44, 1108–1114. doi:10.1086/512818.

Reynolds, L. A., Smith, K. A., Filbey, K. J., Harcus, Y., and James, P. (2016). Commensal-pathogen interactions in the intestinal tract: Lactobacilli promote infection with, and are promoted by, helminth parasites. *Gut Microbes* 5, 522–532. doi:10.4161/gmic.32155.Commensal-pathogen.

Roehe, R., Dewhurst, R. J., Duthie, C. A., Rooke, J. A., McKain, N., Ross, D. W., et al. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* 12, e1005846. doi:10.1371/journal.pgen.1005846.

Shinkai, T., Ueki, T., and Kobayashi, Y. (2010). Detection and identification of rumen bacteria constituting a fibrolytic consortium dominated by Fibrobacter succinogenes. *Anim. Sci. J.* 81, 72–79. doi:10.1111/j.1740-0929.2009.00698.x.

Sun, Y., Su, Y., and Zhu, W. (2016). Microbiome-metabolome responses in the cecum and colon of pig to a high resistant starch diet. *Front. Microbiol.* 7, 1–10. doi:10.3389/fmicb.2016.00779.

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C.-A., Hyslop, J. J., Ross, D. W., et al. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 16, 839. doi:10.1186/s12864-015-2032-0.

Wang, Y., Nan, X., Zhao, Y., Jiang, L., Wang, M., Wang, H., et al. (2021). Rumen microbiome structure and metabolites activity in dairy cows with clinical and subclinical mastitis. *J. Anim. Sci. Biotechnol.* 12, 1–21. doi:10.1186/s40104-020-00543-1.

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J., et al. (2008). Assessment of PLSDA cross validation. *Metabolomics* 4, 81–89. doi:10.1007/s11306-007-0099-6.

Willey, J. M., Sandman, K. M., and Wood, D. H. (2020). “Proteobacteria,” in *Prescott’s Microbiology* (New York), 504–536.

Zeng, H., Guo, C., Sun, D., Seddik, H. E., and Mao, S. (2019). The ruminal microbiome and metabolome alterations associated with diet-induced milk fat depression in dairy cows. *Metabolites* 9. doi:10.3390/metabo9070154.