Norwegian University
of Life Sciences

Master's Thesis 2019 60 ECTS
Faculty of Chemistry, Biotechnology, and Food Science

Understanding rumen function using (meta)genome-guided metaproteomics

Thea Os Andersen
Master of Science, Biotechnology

Understanding rumen function using (meta)genome-guided metaproteomics

Master's Thesis

Thea Os Andersen

Protein engineering and Proteomics Group

Faculty of Chemistry, Biotechnology, and Food Science

Norwegian University of Life Sciences

2019

Acknowledgements

The research presented in this thesis was performed at the Faculty of Chemistry, Biotechnology, and Food Science at the Norwegian University of Life Sciences (NMBU) under the supervision of Assoc. Prof. Phillip B. Pope, Dr. Magnus Øverlie Arntzen and Dr. Live Heldal Hagen.

Firstly, I would like to thank the French National Institute of Agricultural Research (INRA) for providing the samples and for good collaborations, making this thesis possible. I would also like to thank Praveen Kumar and Tim Griffin at the University of Minnesota for allowing and helping us to include and use the Galaxy workflow in this thesis.

I would like to thank my main supervisor, Assoc. Prof. Phillip B. Pope, for introducing me to this project and for adapting the thesis to my fields of interest. I am grateful for your encouragement throughout the process and for sharing of your knowledge. Thank you for including me as an equal in your group and helping me develop myself in the academic field.

Dr. Magnus Øverlie Arntzen and Dr. Live Heldal Hagen, thank you for everything you have taught me and helped me with in the lab, and in regard to databases and data analysis. Magnus, thank you for being calm and patient, and for teaching me the importance of thoroughness. Thank you, Live, for always checking up on me, proof reading and believing in my abilities. Thank you both for your constructive feedback and for pushing me to work hard and forward.

Thank you, Dr. John Christian Gaby, for teaching me what science really is; failing and learning from it. Thank you, Senior Engineer Morten Skaugen, for the help with protein extractions, MS runs and MaxQuant runs.

I would also like to thank the Protein engineering and Proteomics group (PEP), under the leadership of Prof. Vincent Eijsink, for allowing me to write my master thesis in the group. A thanks to everyone at PEP for always having the time to help me around the lab, and for including me in a warm and extremely skillful environment.

I would like to thank all of my fellow master students, new friends and old. Thank you so much for supporting and encouraging words throughout this year, for lunch breaks, beers, though

laughs and tears. I am lucky to have such an ambitious, smart and funny group of friends as you. I wish the best for all of you.

Lastly, I would like to thank my friends and family, my parents, Julianne and Ragnhild, for supporting me in every way you know how. Thank you for believing in me when I don't and for inspiring me to be my best self.

Ås, May 2019
Thea Os Andersen

Abstract

Enteric methane production in ruminants is a major anthropogenic source of methane emissions and contributor to global climate change. The metabolic functions carried out by the rumen microbiome are of scientific and industrial interest, as increased understanding of this complex microbial community can contribute to increased feed efficiency and production of important human food sources (meat and dairy), as well as development of methane mitigation strategies, without compromising livestock.

The microbial community present in the rumen is composed of a dense and complex mixture of anaerobic bacteria, protozoa, archaea, fungi and phages. In close synergetic relationships, the rumen microbiota specializes in degradation and fermentation of complex lignocellulosic biomass into volatile fatty acids (VFAs), which is utilized for host energy metabolism. In addition to VFAs, anaerobic fermentation produces carbon dioxide and hydrogen as byproducts, which subsequently can be converted to methane by archaeal populations known as methanogens. While our knowledge regarding rumen microbiology has been built from studying isolated microorganisms in pure culture, advances in culture-independent "multi-omics" approaches are making increasingly larger contributions to a greater understanding of previously understated environmental microbial communities in the rumen.

In this study, the rumen metaproteome of samples from cows and goats subjected to different diets supplemented with different lipid sources and complex carbohydrates was analyzed. To do this, we utilized a cow-rumen specific protein sequence database consisting of metagenome-assembled genomes (MAGs) and genomes from cultivated rumen bacteria and fungi. We illustrate that database design has a major influence in the identification rate of proteins that are expressed by complex microbial communities. By combining genomic data with metaproteomics, downstream analysis using high accuracy mass spectrometry and advanced proteomics software enabled a detailed portrait of the rumen microbial community and its metabolic functions. Our "multi-omic" analysis reflected a change in rumen microbial composition in cases where cows were subjected to a high starch diet (COS), in addition to an indirect shift in metabolic pathways that suggested lower hydrogen production, which is hypothesized to be connected to the (lower) activity of methanogenic archaea. These observations were supported by lower methane measurements in the host animals feed COS. Moving forward, improvements to database design, which include using sample-specific

metagenomes, will transform our ability to extract maximum understanding into complex rumen microbial populations, their metabolic functions and their interactions with the host. Ultimately, this will create greater opportunities to manipulate the rumen microbiome and to decrease greenhouse gas emissions while ensuring animal health.

Sammendrag

Metanproduksjon hos drøvtyggere er en stor menneskeligskapt kilde til metanutslipp og bidragsyter til globale klimaendringer. De metabolske funksjonene som foregår i mikrobiomet i vomma er av vitenskapelig og industriell interesse, ettersom økt forståelse av dette kompliserte mikrobielle samfunnet kan bidra til økt fôreffektivitet og produksjon av viktige humane kilder til mat (kjøtt og meieriprodukter), i tillegg til utvikling av metanminimerende strategier, uten å gå på bekostning av bærekraftig dyrehold.

Det mikrobielle samfunnet i vomma består av en mangfoldig og kompleks komposisjon av anaerobe bakterier, protozoer, arker, sopp og fager. Gjennom nære, synergistiske forhold spesialiserer mikrobiotaen i vomma seg på degradering og fermentering av kompleks lignocellulosisk biomasse til flyktige fettsyrer (VFA), som kan utnyttes i vertens energimetabolisme. I tillegg til VFA, produserer anaerob fermentering karbondioksid og hydrogen som biprodukter, som videre kan konverteres til metan av arkepopulasjoner kjent som metanogener. Selv om vår kunnskap om vommas mikrobiologi tidligere kun har vært bygget på studier av isolerte mikroorganismer i renkultur, bidrar framgang i kultiveringsuavhengige «multi-omics»-tilnærminger til stadig økte bidrag mot en større forståelse av tidligere underestimerte mikrobielle samfunn tilstede i naturen, slik som i vomma.

I denne studien ble vommas metaproteom fra prøver fra kuer og geiter fôret på ulike dietter tilsatt ulike lipidkilder og komplekse karbohydrater analysert. For å gjøre dette, har vi brukt en kuvom-spesifikk proteinsekvensdatabase bestående av metagenom-assemblerte genom (MAGs) og genom fra kultiverte bakterier og sopp fra vomma. Vi illustrerer at databasedesign har stor innflytelse på identifikasjonsraten av proteinene som utrykkes i komplekse mikrobielle samfunn. Ved å kombinere genomdata med metaproteomikk, kan man ved hjelp av nedstrømsanalyse høy-nøyaktig massespektrometri og avansert protemikk-software, skape et detaljert portrett av det mikrobielle samfunnet i vomma og dets metabolske funksjon. Vår «multi-omics»-analyse reflekterte en endring i mikrobiell sammensetning i tilfeller hvor kuer ble fôret en stivelsesrik diett (COS), i tillegg til en indirekte endring i populasjonen av metanogene arker, som indikerer lavere hydrogenproduksjon, som igjen er foreslått koblet til (lavere) aktivitet hos metanogene arker. Disse observasjonene ble støttet opp under av lavere metanmålinger hos vertsdyrene som var fôret COS. For fremtiden vil forbedringer av databasedesignet, som inkluderer bruken av prøvespesifikke metagenom, omforme vår evne til

å utvinne maksimal forståelse av mikrobielle populasjoner i vomma, deres metabolske funksjoner og deres interaksjoner med verten. Dette vil, igjen, skape større muligheter for å manipulere mikrobiomet i vomma og minske klimagassutslipp, samtidig som man sikrer dyrehelsen.

Abbreviations

A_{595}, A_{750}	Absorbance measured at 595 nm / 750 nm
ACN	Acetonitrile
AmBic	Ammonium bicarbonate
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
CAZymes	Carbohydrate active enzymes
DC	Detergent compatible
DTT	Dithiothreitol
FA	Fatty acids
FDR	False discovery rate
HPLC	High performance liquid chromatography
IAA	Iodoacetamide
LFQ	Label free quantification
mA	Milliampere
Milli-Q	ddH_2O, double distilled water
MS	Mass spectrometer, mass spectra
MS/MS	Tandem mass spectra
Nano LC-MS/MS	Nano liquid chromatography tandem mass spectrometry
PSM	Peptide-spectrum match
PTM	Post transcriptional modification
PUFA	Poly unsaturated fatty acids
PUL	Polysaccharide utilization loci
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
TFA	Trifluoroacetic acid
v/v	Volume/volume
VFA	Volatile fatty acids

Table of content

1 Introduction

1.1 Background

In the last century, human induced greenhouse gas emissions have been one of the leading causes of climate change (Stocker et al., 2013). The Intergovernmental Panel on Climate Change (IPCC) report from 2013 shows that global air temperatures have increased over the last 100 years, as atmospheric concentration of greenhouse gases, such as carbon dioxide (CO_2), nitrous oxide (N_2O) and methane (CH_4) have increased (Stocker et al., 2013). On a global level, greenhouse gas emissions have increased by 75% since 1970 (Edenhofer et al., 2014). CO_2, CH_4 and N_2O contributes to respectively 72%, 19% and 6% of the total global greenhouse gas emissions (Edenhofer et al., 2014; Olivier et al., 2017). Methane has 25 times the global warming potential than that of CO_2 and is therefore an important and significant greenhouse gas (Yvon-Durocher et al., 2014). Methane is emitted through natural gas from fossil fuels, waste decomposition of organic material, and agriculture and livestock production. In 2016, the production of methane from enteric fermentation, primarily from ruminant animals, represented 16% of the total global methane emissions (Olivier et al., 2017). The microbiome in the foregut (rumen) of ruminants, such as beef and dairy cattle, goats, sheep, deer etc., produce methane through anaerobic fermentation of complex carbohydrates (EPA, 2019; Moran, 2005). The amount of methane produced from enteric fermentation in ruminants depends on the individual animal's digestive system and diet (EPA, 2019).

Several studies have been conducted on the mitigation of methane emissions from ruminants using a multitude of different approaches (Beauchemin & McGinn, 2006; Hristov et al., 2013; Roque et al., 2019). A review by Hristov et al. from 2013 presented possible options for methane mitigation including inhibition of methanogenesis, manipulation of the rumen microbiome, addition of lipids to the diet, and exogenous enzymes to mention a few (Hristov et al., 2013). Hristov et al. concluded that while long-term assessments of different mitigation options are not yet well established, it is important that mitigation options are not compromising livestock production.

1.2 Ruminant digestive system

The gastrointestinal tract of ruminant animals consists of four digestive compartments for digestion of consumed lignocellulosic biomass and absorption of nutrients; the rumen, reticulum, omasum and abomasum (**Figure 1.1**). In an adult ruminant animal, the rumen the biggest of the four compartments (Moran, 2005). Initially, ingested feed flows though the rumen and the reticulum, where biomass is broken into smaller pieces through rumination (cud-chewing) biomass. Rumination makes the biomass more susceptible to carbohydrate hydrolysis and further fermentation. As ruminants and most other animals lack carbohydrate active enzymes (CAZymes) in order to degrade plant biomass, the ruminant host depends on the presence of microorganisms in the rumen to produce CAZymes and utilize these carbohydrates (Van Soest, 1994; Wallace et al., 2015). The rumen maintains a stable pH value of about 5.5-7 and a temperature around 39°C (Antanaitis et al., 2016; Moran, 2005), making the rumen optimal for microbial growth and activity. The rumen microbiome is specialized in digestion and fermentation of biomass into nutrients prior to digestion, and includes both Gram-positive and Gram-negative bacteria, protozoa, fungi, phages and archaea. The rumen microbiome and anaerobic digestion of lignocellulosic biomass by the rumen microbiome will be introduced in detail in **Section 1.4**.

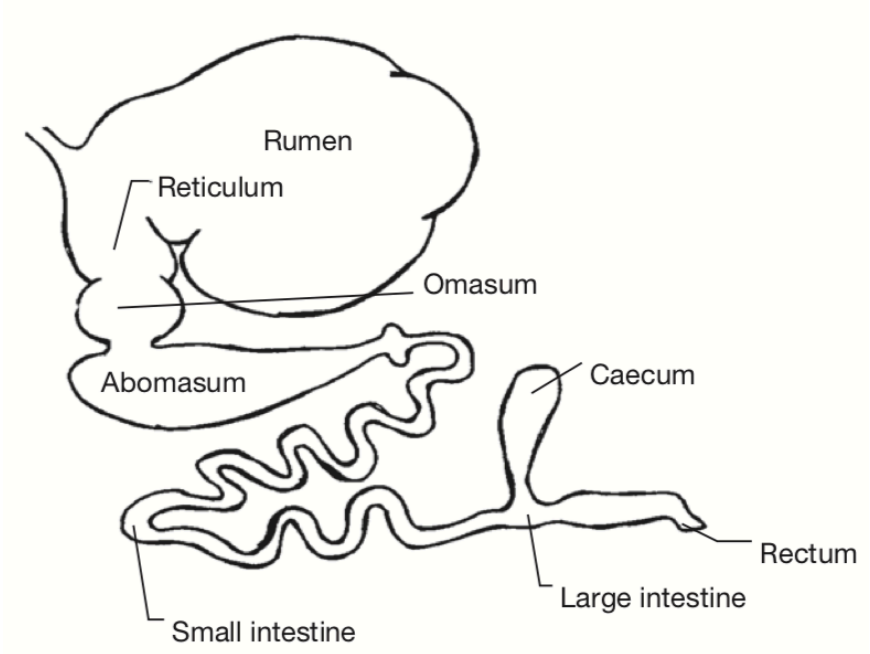


Figure 1.1 The ruminant digestive system. The rumen, reticulum, omasum and abomasum decompose biomass and absorb nutrient. The rumen is the largest of the four compartments and houses a rich microbial community that hydrolyze and ferments complex carbohydrates into nutrient. Figure obtained from Moran (2005).

The rate of passage flow of plant material depends on feed content, its particle size and how efficiently the feed is digested. Digestion efficiency is in turn dependent on abundance of different microorganisms present in the rumen, pH-value and amount of growth limiting or enhancing nutrients (Moran, 2005). After rumination, the feed flows to the omasum. Large folds in the omasum serves as a filter for absorption of water and salt, and further digest the feed (Moran, 2005; Xue et al., 2018). From the omasum, the feed flows to the abomasum, the true stomach. In the abomasum, enzymes and gastric acids further break down feed, much like in the human and mammalian stomach. Just as the rumen houses a large variety of different microorganisms, the reticulum, omasum and abomasum contain microorganisms that contribute to the digestion of feed, even though the microbial composition and digestive contribution in these parts of the ruminant digestive system are currently not well understood (Xue et al., 2018).

1.3. Microbial hydrolysis and fermentation of lignocellulosic biomass

1.3.1 Lignocellulosic biomass

Lignocellulose is the most abundant organic material on earth and consists mainly of polysaccharides, such as cellulose, hemicellulose and pectin, and phenolic compounds (lignin) (**Figure 1.2**) (Moraïs et al., 2012; Scheller & Ulvskov, 2010). Lignocellulose is mainly found in the plant cell wall, where these polymers interact to create a rigid and recalcitrant structure, which is largely due to the crystallinity of cellulose fibers, hydrophobicity of lignin and the lignin-hemicellulose matrix encapsulation of cellulose (Moraïs et al., 2012). The rigidity of lignocellulose makes it challenging to degrade, yet lignocellulose is efficiently digested by microorganisms present in the rumen. Understanding microbial carbohydrate degradation of lignocellulosic biomass in the rumen is of great interest to scientists worldwide as it helps gain insight into the natural carbon cycle, in order to provide the world with renewable sources of energy (Naas et al., 2017; Yue et al., 2013).

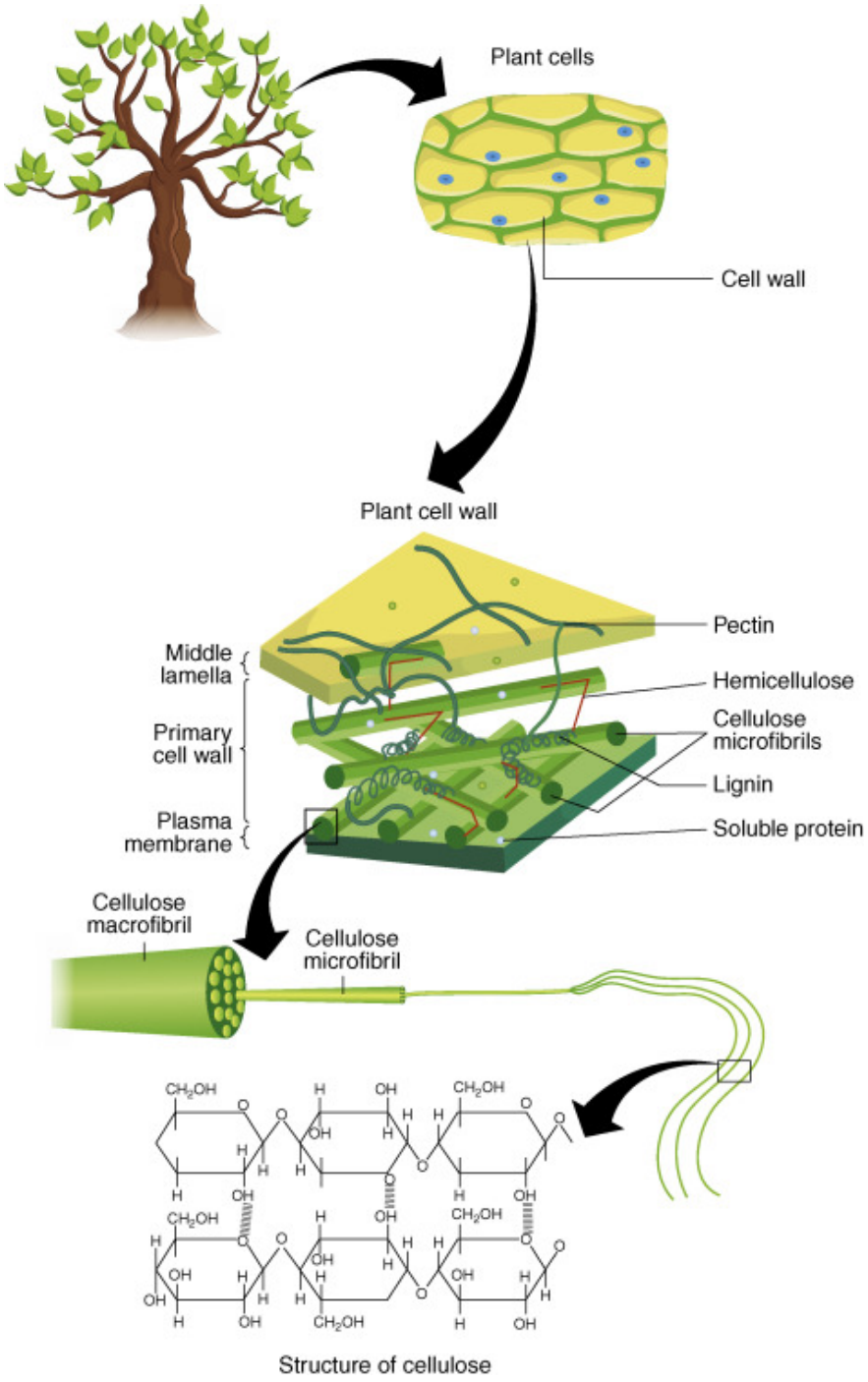


Figure 1.2 Structure of lignocellulose. Lignocellulose is mainly found in the plant cell wall, where polysaccharides (cellulose, hemicellulose, pectin) and phenolic compounds (lignin) create a rigid structure. Cellulose forms cellulose microfibrils consisting of glucose units. Figure obtained from Gupta et al. (2016).

Cellulose is a large, energy rich linear polysaccharide consisting of β-1,4-linked glucose units and compromises 25-50% of lignocellulosic biomass (Naas et al., 2017; Wyman & Yang, 2009). In the plant cell wall cellulose units form rigid crystalline structures due to hydrogen bonds and van der Waals interactions between parallel cellulose chains, called cellulose microfibrils (**Figure 1.2**) (Horn et al., 2012). While cellulose consists only of glucose, hemicellulose consists of many different five- and six-carbon sugar monomers. Hemicellulose is a broad term for other non-cellulosic polysaccharides present in lignocellulosic biomass, compromising xyloglucans, mixed linkage glucans, xylans, mannans, glucomannans and galactomannans (Naas et al., 2017; Scheller & Ulvskov, 2010). Hemicelluloses are generally easier to degrade than celluloses, but complex branching of hemicelluloses contributes to a more recalcitrant structure. To the contrary of cellulose, hemicellulose consists of variety of different sugars that can be difficult to ferment (Horn et al., 2012). Pectin is another polysaccharide component of lignocellulosic biomass and is mainly found in plant cell walls. Pectins are complex polysaccharides composed of homogalacturonans, rhamnogalacturonans and xylogalacturonans, consisting of α-1,4-galacturonic acid (Xiao & Anderson, 2013). In nature, pectins are often metyl-esterified or acetylated, which increases the recalcitrance and complicates enzymatic degradation. However, pectins show promising properties as high-value renewable biomass co-product (Xiao & Anderson, 2013). Lignin is a complex phenolic compound and contributes to the structural rigidity of the plant cell wall, due to cross-linkage with polysaccharides (Naas et al., 2017). The recalcitrant structure of lignocellulose (partly) provided by lignin requires consequently a pretreatment step in industrial degradation of lignocellulose, separating lignin from cellulose and hemicellulose (Naas et al., 2017). Lignin cannot be degraded into sugars, but can be utilized as raw material for making of other aromatic compounds (Wyman & Yang, 2009).

1.3.2 Microbial and enzymatic degradation of carbohydrates

As the ruminant host is not able to degrade lignocellulosic biomass, it is reliant on the symbiotic relationships between microorganisms in the rumen to metabolize lignocellulosic biomass into simpler sugars and starch, which can be utilized for host energy metabolism (Van Soest, 1994; Wallace et al., 2015). The microorganisms in the rumen utilize a vast amount of different CAZymes to break down lignocellulosic biomass into sugars that are ultimately converted via fermentation into short-chain volatile fatty acids (VFAs), such as acetate, butyrate, propionate

and lactate, as well as ethanol and hydrogen gas (Lombard et al., 2013; Marvin-Sikkema et al., 1990). The synthesis and degradation of complex carbohydrates like those mentioned above are controlled by carbohydrate active enzymes called CAZymes (Lombard et al., 2013). In the Carbohydrate-Active Enzyme database (CAZy; http://www.cazy.org) database from 1999 carbohydrate active enzymes and proteins that act on carbohydrates are divided into six different classes based on their mode of action (Lombard et al., 2013). The glycotransferases (GT) class are transferases responsible for assemblage of carbohydrates as they introduce glycoside linkages. Glycoside hydrolases (GH), polysaccharide lyases (PL) and carbohydrate esterases (CE) are all enzyme classes associated with carbohydrate degradation (Lombard et al., 2013). Furthermore, the CAZy database include carbohydrate binding modules (CBMs) that help enzymes target their substrate, and enzymes with auxiliary activity (AA), e.g. lytic polysaccharide monooxygenases (LPMOs) (Lombard et al., 2013; Villares et al., 2017). Through different microbial strategies CAZymes are employed for the degradation of lignocellulosic biomass. The best known mechanisms for microbial degradation of lignocellulose are cellulosomes, free secreted cellulases and polysaccharide utilization loci (PULs) (Naas et al., 2017).

Cellulosomes are cellulolytic multi-enzyme complexes produced by anaerobic cellulolytic microorganisms that were first described in the anaerobic thermophilic soil bacterium *Clostridium* (*Ruminiclostridium) thermocellum* (Moraïs et al., 2012; Schwarz, 2001), but have since been found in the rumen (*Ruminococcus flavefaciens*) (Ding et al., 2001). In addition to their catalytic domain, the cellulases in the cellulosome contain dockerin domains, which can be docked to cohesion domains in the scaffold protein, often bound to the cell surface (Naas et al., 2017). Some microorganism, like *C. thermocellum*, have a CBM on the scaffold protein binding to the cellulose crystalline structure (Moraïs et al., 2012). The mechanism of secreted free enzymes can be observed primarily in aerobic fungi and bacteria but is also present in anaerobic bacteria. A mixture of different cellulases are secreted out of the microbial cells to degrade and utilize cellulose (Naas et al., 2017). Polysaccharide utilization loci (PULs) were first described in the *Bacteroidetes* phylum and consists of gene clusters encoding starch utilizing system (Sus)-like genes, glycoside hydrolases, sugar transporters and regulatory proteins, allowing the bacteria to respond to, bind and utilize a variety of polysaccharides (Naas et al., 2014; Naas et al., 2017). The following sections will introduce the composition of the rumen microbiome and how hydrolyzed carbohydrates can be further fermented into metabolic compounds that are used for ruminant metabolism.

1.4 The rumen microbiome

As mentioned above, the rumen microbiome consists of many obligate and facultative anaerobic species, including bacteria, protozoa, fungi, virus and archaea (Roque et al., 2019; Wallace et al., 2015). The rumen core microbiome consists predominantly of the three bacterial phyla *Bacteroidetes*, *Firmicutes* and *Proteobacteria*, at levels of respectively 33%, 43% and 14%. In addition, less abundant bacterial phyla, such as *Fibrobacteres* and *Spirochaetes* (Petri et al., 2014) also consist of key carbohydrate degrading and hydrogen producing species.

Degradation of cellulose in the rumen is highly associated with the bacterial phyla *Firmicutes* and *Fibrobacteres* (Naas et al., 2014; Petri et al., 2014). Gram-positive *Firmicutes* species such as *Ruminococcus albus* and *Ruminococcus flavefaciens* produce cellulosomes for lignocellulosic degradation in the rumen (Naas et al., 2017). *Fibrobacter succinogenes* is another well studied anaerobic bacteria that produces succinate from degradation of cellulose. While the *Bacteroidetes* phylum is not considered to house cellulolytic populations, many saccharolytic species, including *Bacteriodes* and *Prevotella* species can degrade a wide variety of hemicellulosic and pectin substrates (Naas et al., 2014; Petri et al., 2014). The microorganisms responsible for methane production in the rumen are a group of archaea collectively called methanogens, but very few have been isolated from the rumen. Most of the studied rumen methanogens belongs to the archaeal genera *Methanomicrobium* and *Methanobrevibacter* (Janssen & Kirs, 2008).

In addition to the abovementioned prokaryotes, the rumen microbiome also consists of eukaryotes including protozoa and anaerobic fungi. Although they are important for the rumen function, the contribution of the rumen eukaryotes is not yet fully understood. To date, only one protozoa (Diaz-Viraque et al., 2018) and six anaerobic fungi (Gruninger et al., 2014) are characterized due to technical difficulties associated with their cultivation as well as sequencing and analyzing their eukaryotic genomes. All six anaerobic rumen fungi belongs to the phylum *Neocallimastigomycota* (Gruninger et al., 2014; Marvin-Sikkema et al., 1993), including species such as *Neocallimastix*, *Orpinomyces* and *Piromyces.* Anaerobic fungi, such as *Neocalllimastigomycota*, have specialized organelles called hydrogenosomes that produce H_2, CO_2, acetate and ATP through degradation of lignocellulosic biomass in absence of oxygen by free secreted CAZymes and multi modular enzyme complexes (Gruninger et al., 2014; Marvin-Sikkema et al., 1993).

When studying the rumen microbiome, it is important to keep in mind that less than 50% of the rumen microbial population has been identified with culture-based techniques (Seshadri et al., 2018; Yue et al., 2013), leaving the majority of rumen microbiome and thereby its potential unknown. Modern culture independent techniques, such as whole genome sequencing and -omics techniques, are employed to identify and deeper understand the role of the rumen microbiome. Techniques for studying complex microbial communities, like the rumen microbiome, will be introduced in **Section 1.5**.

1.4.1 Carbohydrate fermentation in the rumen

As previously stated, hydrolysis and fermentation of carbohydrates in the rumen is the result of microbial-driven symbiotic relationships within the rumen microbiome. A simplified overview of carbohydrate hydrolysis and fermentation by the rumen microbiome is showed in **Figure 1.3**. The degradation or hydrolysis of lignocellulosic biomass into simpler sugars by some microorganisms, such as *Ruminococcus*, *Fibrobacter*, *Prevotella* etc., enables other microorganisms to further ferment sugars into hydrogen gas (H_2) and carbon dioxide (CO2), acetate, butyrate, succinate, propionate, lactate and formate (Dijkstra, 1994; Marvin-Sikkema et al., 1993). While many rumen microorganisms can produce lactate, succinate and formate as fermentation intermediates, they seldom accumulate in the rumen and are often converted into acetate, propionate or methane (Figure 1.3). In particular, formate and H_2 can further be utilized in the production of methane (CH_4) via methanoges, whereas VFAs such as acetate, butyrate and propionate are absorbed by the ruminant host and used for energy and growth (Dijkstra, 1994).

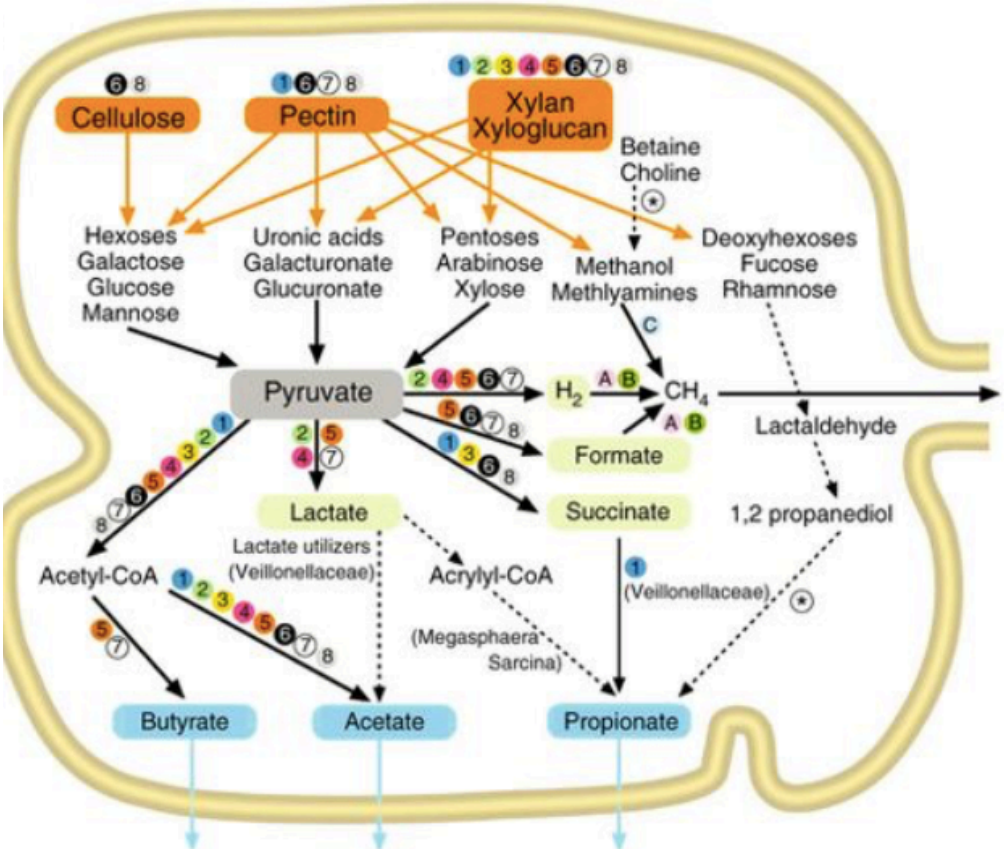


Figure 1.3 Simplified overview of degradation and metabolism of carbohydrates by rumen microorganisms. Numbers and letters represent different bacterial or archaeal groups responsible for conversion of one metabolic compound to another. Numbers for bacterial groups are 1) *Prevotella*, 2) *Clostridiales*, 3) *Bacteroidales*, 4) *Ruminococcaceae*, 5) *Lachnospiraceae*, 6) *Ruminococcus*, 7) *Butyrivibrio*, 8) *Fibrobacter.* Letters for archaeal groups are A) *Methanobrevibacter gottschalkii*, B) *Methanobrevibacter ruminantium,* C) *Methanomassiliicoccales* group 12 sp. This figure shows how this complex composition of different microorganisms is dependent on the syntrophic relationships that exist between species to hydrolyze and ferment carbohydrates into gasses and other simpler compounds for their own metabolism. For the ruminant host, important end products of carbohydrate degradation, such as VFAs, are crucial as they are a major source of energy for the host, used for general metabolism, growth and milk production (Dijkstra, 1994; Moran, 2005). Figure obtained from Seshadri et al. (2018).

The ratio of VFA production is dependent on the type of feed digested. For example, fermentation of structural carbohydrates such as pectin and hemicellulose yields higher amounts of acetate than propionate, the latter of which is an end product of fermentation of sugars and starch (Dijkstra, 1994; Moran, 2005). Hydrogen is released in the first step of formation of acetate, which again can act as an electron donor for production of methane by methanogenic archaea in the rumen (Marvin-Sikkema et al., 1993). Acetate is also important for production of milk fat. In high grain or low fiber diets the production of acetate are lower, which can cause milk fat depression (Fougère & Bernard, 2019; Moran, 2005). The role of

butyrate in the rumen is not fully understood, but studies indicate that butyrate plays a role in regulation of urea kinetics and increase the absorptive capability of VFAs in the epithelium (Agarwal et al., 2015). Butyrate is also metabolized in the liver as ketone bodies, that can be utilized as energy for fatty acid synthesis (Moran, 2005). Propionate production is considered more efficient for the ruminant rather than production of the other VFAs (Dijkstra, 1994; Marvin-Sikkema et al., 1990). This is because production of propionate acts as a hydrogen sink and is thus a competing pathway to methanogenesis.

More recently, non-terrestrial carbohydrates have been trialed to assess their effect on VFA and methane production. A 2019 study conducted by Roque et al. examined the methane mitigation effects of the macroalgae *Asparagopsis taxiformis* and discovered that reduction of methane increased the propionate:acetate ratio, suggesting that hydrogen is (somewhat) redirected or redistributed to propionate (Roque et al., 2019). Production of propionate and butyrate is positively associated with increased milk yield, yet decreased acetate ratios, as mentioned, can impact milk fat production (Moran, 2005). Propionate can, in addition, be used as a precursor for glucose synthesis (Dijkstra, 1994). Inhibiting methanogenesis in this manner relieves H_2 for VFA production, more favorable for the ruminant and potentially increase feed efficiency. On the contrary, studies have in some cases shown that inhibition of methanogenesis can affect the production of VFAs either by decreasing or increasing their production, thus reducing feed efficiency (Machado et al., 2016; Roque et al., 2019). In negative instances, an increase in VFAs can lead to accumulation of lactic acid, that will decrease the pH in the rumen and create a hostile environment for microorganisms, and in worst case scenarios cause death in ruminants (Annison et al., 2007).

Methane production has been estimated to amount to a 2-12% energy loss for the animal (Tapio et al., 2017), as hydrogen is converted to methane instead of other VFAs for host metabolism utilization. In the rumen, hydrogen formed from fermentation can be converted into methane by reducing CO_2 through the hydrogenotrophic methanogenesis pathway by methanogenic archaea (Wallace et al., 2015). Conversion of H_2 to methane is an exergonic reaction and thereby a favorable reaction over the production of other H_2 sink products, such as VFAs (Marvin-Sikkema et al., 1990). As methanogens act as a H_2 sink, they help keep the H_2-partial pressure low, in order to ferment compounds further into H_2, CO_2 and acetate. An increase in high hydrogen partial pressure, which is critical to maintain homeostasis in many fermentative Gram-positives (e.g. *Ruminicoccus* sp.) The production of the fermentation end products is

therefore highly dependent on syntrophic relationships between H_2-producing species and hconsuming species (Marvin-Sikkema et al., 1990). While most of our knowledge in ruminal digestion has been built from prokaryotes, in 1990 Marvin-Sikkema et al. studied the effects of cellulose degradation by anaerobic fungi in the presence of hydrogen consuming microorganism, including some methanogens. They concluded that the presence of methanogens caused a production shift in the anaerobic fungi, producing more H_2 that other electron sink products, like VFAs, and thereby favoring production of ATP and acetate, which again had an effect on the energy metabolism and growth of the fungus (Marvin-Sikkema et al., 1990).

1.4.2 Dietary lipid protection in the rumen

In recent decades, lipid supplements to ruminant diet have been used to improve energy intake in high-producing dairy cows (Fougère & Bernard, 2019). Lipids are useful as supplements in ruminant diets, as the high caloric value of lipids can be a powerful tool in overcoming limitations in energy supplies in high-yielding ruminants (De Beni Arrigoni et al., 2016). Additionally, lipids can be used to manipulate the digestion and absorption of nutrient in the rumen and limit ruminal acidosis and depressed milk fat content resulting from high carbohydrate and low fiber diets. Ruminant diets with supplemented lipid can also alter the proportion of particular fatty acids (FAs) in meat or milk fat to be more desirable for the food industry or for the consumption by humans (De Beni Arrigoni et al., 2016). Dietary fatty acids usually constitute less that 3% of the ruminant diets (Chilliard, 1993), and are often found in forage, grains and seeds, such as soybean oil, corn oil and palm oil (Fougère & Bernard, 2019). In common for lipids from forages supplemented in ruminant diets are high proportions of polyunsaturated fatty acids (PUFAs), particularly linoleic or linolenic acids, which are essential fatty acids in the diet as they cannot be synthesized by the ruminant or by humans (Owens & Basalan, 2016) and constitute an important part of a healthy human diet. Increased PUFA concentrations are also related to increased fertility and better reproductive performances of ruminants (Gadeyne et al., 2017).

However, improving the content of PUFA in ruminant products, such as milk and meat, have proven to be challenging as due to microbial saturation of dietary PUFAs in the rumen. Through the process of biohydrogenation (saturation), PUFAs are converted into hydrogenated products by biohydrogenating bacteria in the rumen, such as *Butyrivibrio fibrisolvens*, before absorption

in the small intestine (Lourenço et al., 2010). In biohydrogenation processes, hydrogen in the rumen is used to saturate PUFA, thereby affecting the rumen fermentation and production of other hydrogen sink products in the rumen, such as methane and VFA that can be utilized for host energy metabolism. The development of new protection technologies in order to overcome biohydrogenation of PUFAs are of great scientific interest and importance in order to provide sustainable livestock production and healthy animals (Fougère & Bernard, 2019; Gadeyne et al., 2017).

1.5 Studying complex microbial communities

Representing 50-78% of the biomass of the earth (Heyer et al., 2017), microorganism exists in every environment, forming symbiotic relationships in complex microbial communities. Understanding the rumen microbiome and its functions can provide greater insight into carbohydrate degradation and methane production, moving towards new and more sustainable solutions for the globally increasing food and energy demands. Yet, knowledge of the rumen microbiome and function are substantially determined by the cultivation eligible part of the microbiome, hence representing only a fraction of the complete rumen microbiome. Utilization of modern cultivation-independent techniques have helped overcome limitations of traditional culture dependent methods, towards an increased understanding of complex systems like the rumen microbiome.

Over the past decades, utilization of cultivation-independent techniques has revealed several novel microorganisms that play a key role in degradation of carbohydrates (Hess et al., 2011; Stewart et al., 2018). The microbial compositions of the rumen can be determined with polymerase chain reaction (PCR) amplification of conserved marker genes. The most widely used marker gene is the 16S ribosomal RNA unit, present in all prokaryotes, together with its eukaryotic equal, 18S ribosomal RNA unit (Klindworth et al., 2013). Several renowned scientists, such as Kary Mullis and Frederick Sanger, the scientists behind respectively PCR (Mullis & Faloona, 1987) and Sanger sequencing (Sanger & Coulson, 1975), have taken advantage of the presence of these conserved marker genes for amplification and sequencing whole genomes of microorganism. As traditional fingerprinting methods, like the abovementioned, are both low-throughput and time consuming, a new generation of culture-independent techniques has emerged in the past decade, reducing cost and at the same time increasing the speed and throughput of sequencing data (Vanwonterghem et al., 2014). These

high throughput sequencing techniques enables the investigation and analysis of a large number of samples that differ in space and/ or time.

By combining high-throughput sequencing techniques with standard laboratory methods, like microscopy or isotope labeling, it is now possible to determine the composition and function of complex microbial communities (Vanwonterghem et al., 2014). Biological studies with aim to characterize and quantify composition and functions are often termed "-omics" approaches, and compromises the study of genes (genomics), transcripts (transcriptomics), proteins (proteomics) and metabolites (metabolomics) (Heyer et al., 2017) (Summarized in **Figure 1.4**). The term 'metagenomics' was first introduced by Jo Handelsman et al. in 1998, which frames the approach in which genomes from multiple microorganisms in a specified environment are analyzed through extraction and cloning of DNA (Handelsman et al., 1998). In combination with modern high throughput sequence techniques, analysis of the metagenomic contribution of a complex microbial communities provide information about phylogenic diversity that would be impossible to achieve with culture dependent methods. In addition to phylogenic understanding, the metagenome can provide knowledge regarding potential functions of the microbial community.

Development of high-throughput sequencing techniques in the past decades have enabled the analysis of not only the metagenome, but also the study of gene expression termed metatranscriptomics (Bashiardes et al., 2016). Metatranscriptomic analysis aims to elucidate annotated genes from metagenomic analysis to understand which genes are transcribed and to what the extent they are transcribed, thus complementing phylogenetic information with information about functional properties of transcriptionally active populations (Bashiardes et al., 2016). As with metatranscriptomics, the study of the metaproteome (i.e. metaproteomics) provides insight into the functional dimension of the microbial community. Protein expression reflects the microbial activities in a given ecosystem at a given time (Wilmes & Bond, 2006) and metaproteomics are therefore central in the understanding of rumen functions. Moreover, metaproteomic analysis has in addition the advantage of providing information about post transcriptional modifications (PTMs) and/ or regulations of gene expression, contrary to metatranscriptomics. As metaproteomics has been essential in the work described in this these, more details on metaproteomics approaches are described in **Section 1.5.1**.

The metabolome of a cell is defined as the total collection of metabolites, small molecules produced by the cell contributing to the biochemical activity (Patti et al., 2012). Unlike genes and proteins that are subjected to regulations and modifications, metabolites serve as direct signatures of biochemical activity (Patti et al., 2012). This taken into consideration, metabolomics or metametabolomics can, in combination with metagenomics or metaproteomics, provide further information about cellular processes and microbial activities in complex microbial communities. In summary, different -omics approaches can be utilized to gain information regarding both what microorganisms are present in a microbial community, and maybe more importantly, provide crucial information about their functions, interactions and microbial activity.

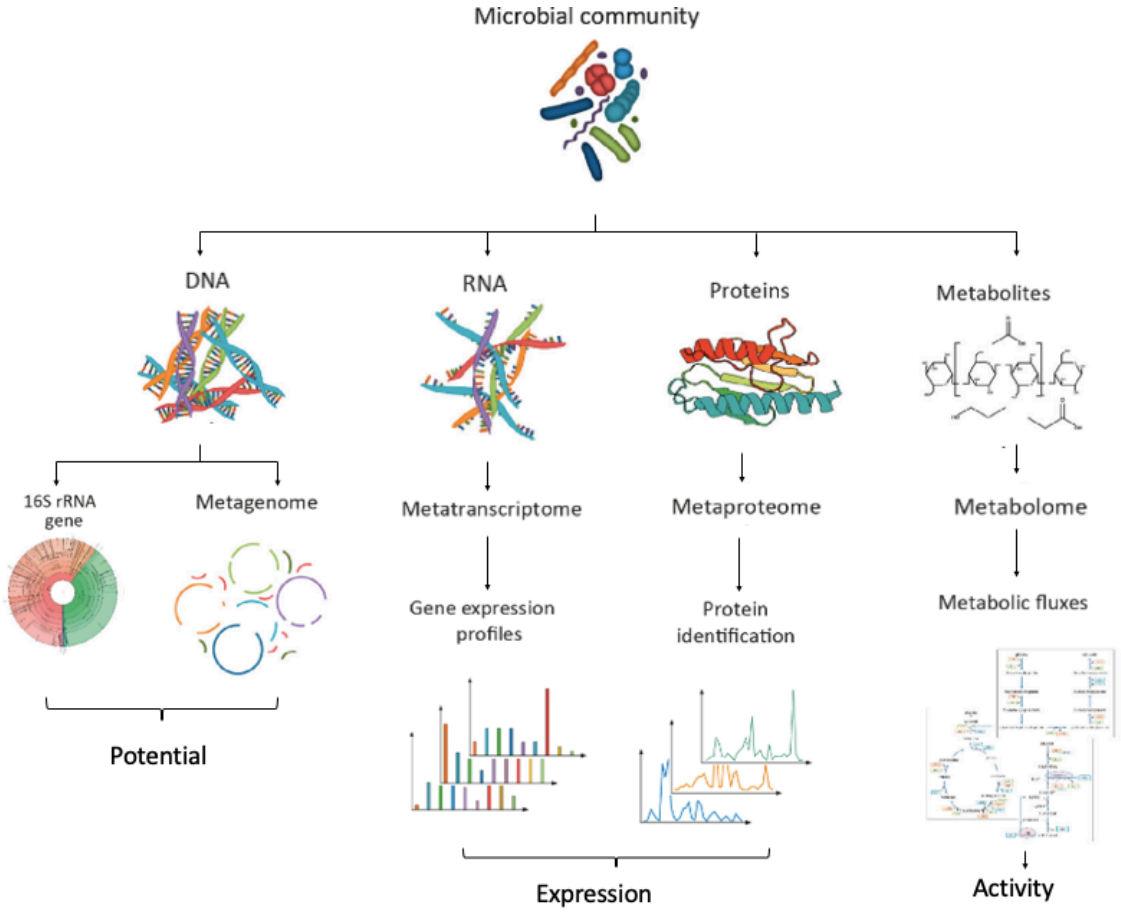


Figure 1.4. Meta-omics approaches provide insight into the potential and the functions of the rumen microbiome. Figure obtained from (Vanwonterghem et al., 2014) with modification.

1.5.1 Metaproteomics

A lot of influential and important discoveries in molecular biology have originated in the seeking of differences between biological systems in various functional states (Cox & Mann, 2007). Different conditional states in cells, e.g. sick or heathy, untreated or treated with stimuli, and protein abundance differences between pathways, can be induced to alter the microbial composition and activity to examine cellular pathways or processes. As proteins carry out most functions in the cell (Heyer et al., 2017), the induction of different functional states in the cell can be examined at the protein level through metaproteomic analysis. The term proteome, first described by Marc Wilkins in 1996, is the entire protein complement expressed by a genome, or by a cell or tissue type (Wilkins et al., 1996). Wilkins also described the proteome as different from the genome, as the genome in an organism are more constant while the proteome can rapidly change in response to altered environments even though the proteome is a direct product of the genome (Wilkins et al., 1996). Alternative gene splicing and PTMs can cause the number of expressed protein-species to exceed the number of genes in an organism (Thiede et al., 2013; Wilkins et al., 1996). About a decade later, metaproteomics was established as the study of the proteome from multiple organisms, enabling the understanding of functional gene expression in microbial communities (Wilmes & Bond, 2006).

1.5.1.1 (Meta)proteomics workflow and mass spectrometry analysis

Proteomics are applicable to both prokaryotic and eukaryotic cells and from a variety of materials, body fluids or whole tissues (Cox & Mann, 2011). Mass spectrometry (MS)-based proteomics has become the *modus operandi* for proteomic analysis today and is also the most commonly applied approach in metaproteomics. In metaproteomic sample preparation, proteins from environmental samples can be analyzed as intact proteins, an approach called top-down proteomics, or as digested peptides, in bottom-up proteomics (Aebersold & Mann, 2016). In the top-down approach, intact proteins (including modifications) are measured and capable of yielding precise identification (Aebersold & Mann, 2016). However, because of difficulties regarding ionization and solubility of proteins, multiple change states per protein, and thus detection of whole proteins is limited and requires both high separation of protein upstream (High Performance Liquid Chromatography (HPLC)) and extreme high-resolution MS (Resing & Ahn, 2005). As the bottom-up proteomics workflow proteins is more manageable, due to the generation of peptides 5-30 amino acids long and having basic C-termini (due to trypsin), it is

thus the most widespread proteomic workflow (**Figure 1.5**)(Cox & Mann, 2007; Cox & Mann, 2011; Heyer et al., 2017).

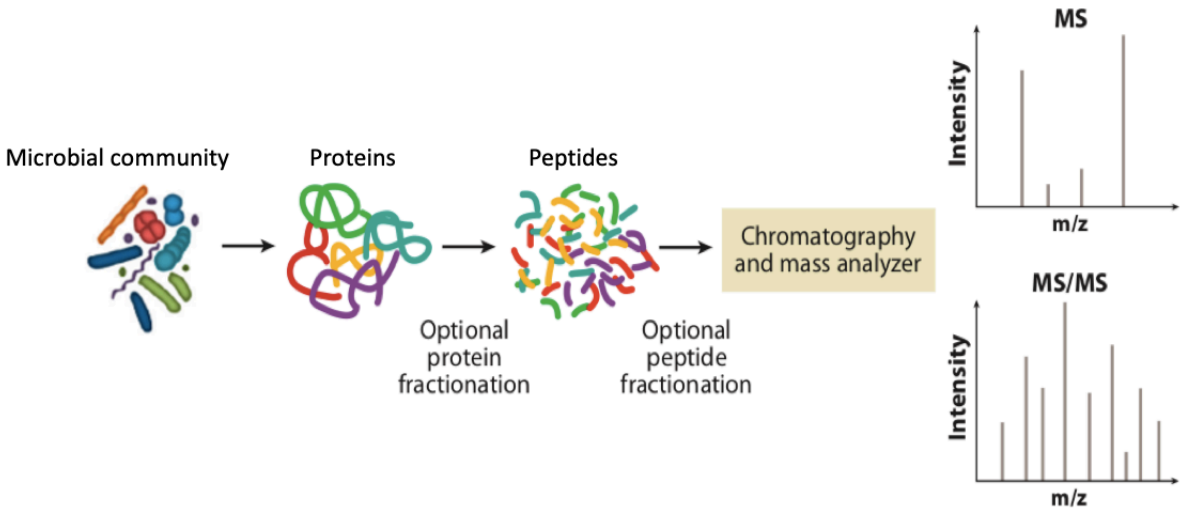


Figure 1.5. Overview of the generic bottom up proteomics workflow. The proteome can be quantified and identified through protein separation in chromatography and mass spectrometry (MS and MS/MS) analysis. See mail text for details about various techniques. Figure obtained from Cox og Mann (2011) with modifications.

The typical (meta)proteomic workflow start with the protein extraction from the source. Protein extraction for metaproteomics from environmental samples is not trivial, as microbial communities consists of a variety of different microorganisms. In order to perform meaningful analyses, it is crucial that the proteins extracted from the samples reflects the present microorganisms as correctly and impartial as possible. Therefore, a number of methods combining chemical and mechanical cell lysis methods have been applied over the last few years (Keiblinger et al., 2016). Most current protocols include the use of a lysis buffer with moderate concentrations of detergent combined with heating and sonication or mechanical cell disruption (Keiblinger et al., 2016). After extraction proteins are either fractionated on one-dimensional gel electrophoresis, i.e. sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE), followed by in-gel digestion, or directly (in-solution) digested into peptides by a sequence-specific enzyme, i.e. trypsin (Cox & Mann, 2011). Peptides can be further separated based on their hydrophobicity using C_{18} columns, with a gradient of aqueous and organic solvents in high-performance liquid chromatography (HPLC) (**Figure 1.5**) (Cox & Mann, 2011; Cravatt et al., 2007). As mentioned above for bottom-up proteomics, peptides are easier to separate and analyze, and is also beneficial for MS analysis, as MS is more sensitive for low

molecular-weight molecules (Cox & Mann, 2011). The duration of the HPLC-gradient can be selected based on the sample (peptide's) complexity, reflecting the complexity of the original microbial community. Depending on the level of fractionating at the protein level, gradients from 30 minutes and up to four hours are commonly applied.

Peptides are ionized in the source of the MS either by aerosol generation and droplet liquid evaporation in electrospray methods, or from peptides embedded in a solid matrix in matrix-assisted laser desorption/ ionization (MALDI). The peptides enter the gas phase and their mass-to-charge (m/z) ratio can be analyzed (**Figure 1.5**, top spectrum). Selected peptides can then be selected for fragmentation to generate tandem MS (MS/MS) spectra (**Figure 1.5**, lower spectrum) (Aebersold & Mann, 2016; Cox & Mann, 2011), that are used to deduce the peptide's amino acid sequence. Additionally, the intensity of the peptide signal reflects the abundance of the ion (Cravatt et al., 2007). State-of-the-art ion-trap MS, such as Orbitrap analyzers, generate mass spectra with high resolutions and with high mass accuracy, while maintaining operation speed, allowing vastly complex mixtures of peptides to be analyzed (Cox & Mann, 2011; Cravatt et al., 2007). The MS can be tuned to select a preset number of peptides, an approach referred to as Top N, when N often varies between 5 and 20 depending on the speed of the MS (Cox & Mann, 2011).

1.5.1.2 Protein identification and quantification

MS raw data contains information about a peptide's m/z ratio and intensity, and this can be used to identify and quantify specific peptides with proteomics software, such as MaxQuant (Tyanova et al., 2016a). MaxQuant scans though the tandem MS raw data and applies its own peptide database search engine, Andromeda, that will return a list of proteins that most likely are present in the samples (Cox & Mann, 2008; Cox & Mann, 2011; Tyanova et al., 2016a). Andromeda uses probability calculations for scoring peptide-spectrum matches (PSMs), that includes user specified allowed variable and/ or fixed protein and peptide modifications, and the digestive enzyme used for protein cleavage, in addition to a user specified protein sequence database (Cox et al., 2011). MaxQuant claims to enable accurate protein quantification, regardless of labeling and fractionating of samples, by using the MaxLFQ workflow. Briefly, the label free quantification (LFQ) intensity of a protein is the relative protein quantification normalized across all samples (Cox et al., 2011).

In order to achieve as accurate protein identification and quantification as possible, strict control of false identifications is necessary. The rate of false positives allowed in a dataset is referred to as the false discovery rate (FDR), and has a default value of 0.01, or 1%, in MaxQuant (Tyanova et al., 2016a). To control the number of false positives, the most common strategy is the target-decoy approach. To achieve this MaxQuant utilize a reversed database, based on the original database but with the reversed sequences of all proteins in the database (Tyanova et al., 2016a). Identified hits to the reversed database will be accepted in the final results as long as this fraction remains under the set (1%) FDR. These hits can be removed before downstream analysis. Proteins with similar primary sequence may share identified peptides. These are combined into protein groups, to avoid overestimation of identifications. As a peptide can only contribute to one protein group's score and quantitative value, peptides that are common to more than one protein group are called razor peptides and will be assigned to the protein group with the largest number of identified peptides (Tyanova et al., 2016a).

The selection of a suitable protein sequence database is essential when it comes to the number of identified proteins and the accuracy of quantification. Optimal databases are sample specific, and include proteins that are present in the sample and detectable on MS (Heyer et al., 2017). For many environmental samples, the exact microbial composition and abundance is unknown. How well the database fits the samples is therefore an important condition for strong FDR estimation (Heyer et al., 2017). The size of the protein sequence database can also affect the FDR; large public repository databases, such as the complete UniProt bacterial section (58 million protein entries) will, due to its enormous size, give few significant hits. The reason for this is the target-decoy strategy itself, as the reversed sequences get a too high score and thus the 1% FDR limit will be reached very fast.

The final step in the (meta)proteomic analysis is the biological interpretation of the acquired protein identifications and abundances. Perseus is a computational platform designed for the analysis of complex proteomics data and combines statistical power with a user-friendly workflow for both the trained bioinformatician and the biochemist (Tyanova et al., 2016b). Perseus can be used to create visual output of data analysis and run statistical test on the data set. The Perseus matrix displays the abundance values of proteins in the biological samples/ replicates of the data set and allows the import of public annotation databases, i.e. Pfam and InterPro, to identify protein function, protein-protein interactions and metabolic pathway annotations (Tyanova et al., 2016b).

1.6 Aim of study

Methane emissions from enteric methanogenesis in ruminants represent 16% of the global methane emissions. Methane is produced as a biproduct in carbohydrate fermentation, carried out by the complex microbial community in the rumen and their symbiotic relationships. Multiple studies have been conducted on the mitigation of methane production from ruminants. In addition to methane, the rumen microbial community ferment structural carbohydrates into volatile fatty acids that can be absorbed by the host and used for growth. The rate of production of methane and volatile fatty acids by the rumen microbial community is highly dependent on the feed digested by the ruminant.

Modern culture-independent techniques, such as sequencing and omics-techniques, have helped gain insight into the rumen microbial community and its function. Metaproteomics represent one such technique that can provide crucial information about protein expression by rumen microorganisms and their abundance in a complex microbial community. Metaproteomic analysis is dependent on high resolution MS and strict control of the FDR to provide accurate protein quantification and identification with proteomics software, such as MaxQuant and Perseus.

In this study 12 Holstein cows and 12 Alpine goats were fed four different feed supplemented with different lipids. These animals represent a larger study conducted at the French National Institute of Agricultural Research (INRA) that aimed to investigate lipid additives in ruminant milk fat secretion and composition (Fougère et al., 2018). In this thesis, we aimed to determine the effects of these dietary interventions on ruminal fermentation and methane production. To do this, we combined obtained metadata measurements for methane and VFAs and metagenome guided-metaproteomic analysis, which was specifically applied to rumen samples to quantify and identify proteins connected to metabolic pathways, such as methane production and VFA production. The application of metaomics techniques can better the understanding of the rumen microbial composition and its metabolic potential and activity. Insight into carbohydrate fermentation as well as methane and volatile fatty acid production, can be utilized to manipulate the rumen microbiome to reduce methane emissions, while securing feed efficiency and a sustainable livestock production.

2 Materials

2.1 Lab equipment

Lab equipment used for the experimental part of the thesis are listed with their respective supplier and catalog number.

PRODUCT	SUPPLIER	CATALOG NUMBER
11 mm Snap Ring Cap	VWR, Pennsylvania, USA	548-0016
Acclaim™ PepMap™ 100 C18 LC Column	Thermo Fischer Scientific, Massachusetts, USA	
BioTek™ Synergy™ H4 Hybrid Microplate Reader	Thermo Fischer Scientific, Massachusetts, USA	
Blue caps for FastPrep® Tubes	MP Biochemicals, Ohio, USA	5065-005
Eppendorf Safe-Lock Tubes (PCR clean)	Eppendorf, Hamburg, Germany 0030120094	0030120094
Eppendorf® epT.I.P.S volume range 0.1-10 µL	Sigma-Aldrich, Missouri, USA	Z640387
Eppendorf® epT.I.P.S volume range 2-200 µL	Sigma-Aldrich, Missouri, USa	Z640336
Eppendorf® epT.I.P.S volume range 50-1000 µL	Sigma-Aldrich, Missouri, USA	Z640433
FastPrep-24™ Classic Grinder	MP Biochemicals, Ohio, USA	SKU116004500
FastPrep® Tubes	MP Biochemicals, Ohio, USA	5076-200

Glass beads, acid washed, ≤ 106 µm, 500g	Sigma-Aldrich, Missouri, USA	G4649-500G
Mini-PROTEAN® Tetra Cell	Bio-Rad, California, USA	
Mini-PROTEAN® TGX Stain-Free™ Gels (Any kD, 10 well comb, 30 µl)	Bio-Rad, California, USA	4561023
PowerPac™ Basic Power Supply	Bio-Rad, California, USA	
Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer	Thermo Fisher Scientific, Massachusetts, USA	IQLAAEGAAPFALGMAZR
ThermoMixer®	Eppendorf, Hamburg, Germany	
UltiMate™3000 RSLCnano System	Thermo Fisher Scientific, Massachusetts, USA	ULTIM3000RSLCNANO
ZipTips® Pipette Tips	Merck-Millipore, Massachusetts, USA	Z720070
0.3 ml PP Snap Ring Micro-Vial	VWR, Pennsylvania, USA	548-0120

2.1.2 General lab equipment

General lab equipment used in the experimental part of the thesis are listed with its respective supplier.

PRODUCT	**SUPPLIER**	**CATALOG NUMBER**
Automatic pipettes	Thermo Fischer Scientific, Massachusetts, USA	
Axygen® 1.5 ml MaxyClear Snaplock Microcentrifuge tube	Corning, New York, USA	MCT-150-C

Branson 3510 Ultrasonic Cleaner	Marshall Scientific, New Hampshire, USA	BR-UC35
Concentrator plus complete system, Vacuum concentrator	Eppendorf, Hamburg, Germany	5305000304
ddH2O, Milli-Q® Reference Water Purification System (0,22 µm filter)	Merch-Millipore, Massachusetts, USA	C79625
Duran® Glass flasks	Shcott, Wertheim, Germany	
Eppendorf BioPhotometer® D30	Eppendorf, Hamburg, Germany	6133000001
Eppendorf® Centrifuge 5418R (4°C)	Sigma-Aldrich, Missouri, USA	EP5401000137
Falcon tubes, 15 ml & 50 ml	Greiner tubes, Sigma-Aldrich, Missori, USA	
Freezer (-20°C)	Bosch, Stuttgart, Germany	
Freezer (-80°C), Innova® C585 Chest Freezer, New Brunswick	MG Scientific, Wisconsin, USA	
Freezer (-80°C), V.I.P.® Series -86°C Ultra Low Freezers	Sanyo, Osaka, Japan	
Heraeus™ Pico™ 21 Microcentrifuge, Centrifuge (22°C)	Thermo Fischer Scientific, Massachusetts, USA 75002553	75002553
IKA® HS 260 Basic Shaker	Thermo Fischer Scientific, Massachusetts, USA	Z341843
Magnetic stirrer, IKA® RCT basic IKAMAG™ Safety Control	Sigma-Aldrich, Missouri, USA	Z645060
MS2 Minishaker IKA® Vortex	Fischer Scientific, New Hampshire, USA	12819435

Nitrile gloves	VWR, Pennsylvania, USA
Quintix® Weight	Santorius, Göttingen, Germany
Refrigerator (4°C)	Bosch, Stuttgart, Germany
Stainless steel surgical blade	Swann-Morton Limited, Sheffield, UK
Water bath (99°C) (5L, 230V)	VWR, Pennsylvania, USA

2.2 Chemicals, manufactured reagents and kits

Chemicals, premade buffers and reagents, and kits are listed with their respective supplier and catalog number.

2.2.1 Chemicals

CHEMICAL	SUPPLIER	CATALOG NUMBER
2-propanol, 2 L	Honeywell, North Carolina, USA	278475
Acetic acid, 100%, 2,5 L	Merck-Millipore, Massachusetts, USA	1.00063.2500
Acetonitrile (I), CHROMASOLV™ LC-MS grade, 1 L	Honeywell, North Carolina, USA	34967-1L
Albumin bovine Cohn Fraction V (BSA), 25 g	Koch-Light Laboratories Ltd., Suffolk, UK	0143-01
Ammonium bicarbonate (AmBic), 500 g	Sigma-Aldrich, Missouri, USA	09830-500G
Coomassie Brilliant Blue R-250, 10 g	Bio-Rad, California, USA	161-0400
Dithiothreitol (DTT)	Sigma-Aldrich, Missouri, USA	D0632-25G

EMSURE® Hydrochloric acid, 32%, 2,5 L	Merck-Millipore, Massachusetts, USA	1.00319.2500
Ethanol absolute, 5 L	VWR, Pennsylvania, USA	20821.365
Iodoacetamine (IAA), 5 g	Sigma-Aldrich, Missouri, USA	I1149
Sodium dodecyl sulfate (SDS), 1 kg	PanReac Applichem ITW Reagents, Darmstadt, Germany	A2572, 1000
Trifluoroacetic acid (TFA), 100%, HiPerSolv CHROMOANORM®, LC-MS grade, 1 L	VWR, Pennsylvania, USA	85049.001
Trizma® base, 1 kg	Sigma-Aldrich, Missouri, USA	T1503

2.2.2 Manufactured buffers, reagents and kits

REAGENT	**SUPPLIER**	**CATALOG NUMBER**
10 x Tris/Glycine/SDS Buffer (TGS), 5 L	Bio-Rad, California, USA	161-0772
BenchMark™, Protein Ladder, 250 µl	Thermo Fisher Scientific, Massachusetts, USA	10747-012
DC™ Protein Assay	Bio-Rad, California, USA	5000111
Novex™ NuPAGE™ LDS Sample Buffer (4X), 250 ml	Thermo Fisher Scientific, Massachusetts, USA	NP0008
NuPAGE™ Sample Reducing Agent (10X), 10 ml	Thermo Fisher Scientific, Massachusetts, USA	NP0009
Protein Assay Dye Reagent Concentrate, 450 ml	Bio-Rad, California, USA	500-0006

Trypsin Porcine, Sequencing grade modified 20 µg	Promega, Wisconsin, USA	V511A
Trypsin resuspension buffer, 1 ml	Promega, Wisconsin, USA	V542A

2.3 Buffers

Protocols for buffers used in this thesis are listed below.

TRIS-HCl 1M 60 ml

- 7,266g Trizma® base was weighed and dissolved in 20 ml Milli-Q.
- pH was adjusted with 1 M HCl until pH = 8.
- Milli-Q was added to reach total volume (60 ml)

LYSIS BUFFER (3X)

1 x	***3 x***
10 mM Dithiothreitol (DTT)	30 mM Dithiothreitol (DTT)
50 mM Tris-HCl (pH=8)	150 mM Tris-HCl (pH=8)
0.1 % Triton X-100	0.3% Triton X-100
4 % SDS	12% SDS

STAIN SOLUTION

25% Isopropanol
10% Acetic acid
0.05% Coomassie Brilliant Blue R-250

DESTAIN SOLUTION

25% Isopropanol
10% Acetic acid

DTT SOLUTION
10 µl 1M Dithiothreitol (DTT)
100 µl 1 Ammionium bicarbonate (AmBic)
890 µl Milli-Q

IAA SOLUTION
10 mg Iodoacetamide (IAA)
100 µl 1M Ammonium bicarbonate (AmBic)
900 µl Milli-Q

TRYPSIN BUFFER
25 µl 1M Ammonium bicarbonate (AmBic)
100µl 100% ACN
875 µl Milli-Q

TRYPSIN SOLUTION
5 µl 500ng/µl Trypsin (frozen at -80°C)
245 µl Trypsin buffer

Solvents in HPLC:

- Solvent A: 0.1% formic acid (v/v) (in water)
- Solvent B: 80% ACN, 0.08% formic acid (v/v) (in water).

2.4 Software tools

NAME	PURPOSE	SUPPLIER	REFERENCE
Xcalibur™, version 3.1.66.10	MS-data acquisition	Thermo Fisher Scientific, Massachusetts, USA	
MaxQuant, version 1.6.3.3	MS-data analysis		(Tyanova et al., 2016a)
Perseus, version 1.6.1.1	Protein quantification and identification		(Tyanova et al., 2016b)
InterProScan, version 5.32-71.0	Functional protein annotation		(Fraser, 2017)

3 Methods

3.1 Sampling

Samples analyzed in this thesis origin from an animal feeding trial conducted at the French National Institute for Agriculture (INRA), as described in **Section 1.6** and in Fougère et al. (2018). Here, Holstein cows and Alpine goat were used in this experiment and fed four different feeds in a 4 x 4 Latin square design (**Table 3.1**) with 28-days experimental periods from February to July 2016 (Fougère et al., 2018). All animals were non-pregnant, multiparous, and at lactation stage of 86 ± 24,9 and 61 ± 1,8 days-in-milk for cows and goats respectively at the time of the experiment (Fougère et al., 2018; Vijayakumar et al., 2017). Four groups, each constituting of 4 cows and 4 goats, were balanced with regards to days-in-milk, milk production, milk fat and milk content (Fougère & Bernard, 2019). Each group was randomly assigned to the one of four diets. The animals were provided grass hay with concentrates supplemented with no added lipid; called **CTL** (control), supplemented with corn oil and wheat starch; called **COS**, supplemented with hydrogenated palm oil; called **HPO**, or supplemented with marine algae powder from the oil-rich microalgae *Schizochytrium* sp.; called **MAP** (**Table 3.2**) (Fougère et al., 2018). All animals were fed the CTL diet for a 16 days adaptation period prior to the experiment start (Fougère et al., 2018). In the experimental periods, all animals were fed the concentrates twice daily, at 08.30 and at 16.00, together with hay. Excess concentrate and hay were weighed daily to adjust the amount of feed given the following day to maintain a forage:concentrate ratio of 45:55 (Fougère et al., 2018). Additionally, the animals had constant water supply ad libitum (Fougère et al., 2018). More detailed information regarding feeding and composition of concentrates is described in Fougère et al. (2018) and in Fougère & Bernard (2019).

Table 3.1 Experimental design. Distribution of diets fed to the animals in each period, according to the 4 x 4 latin square design. Rumen fluid samples were collected in each period and sent to NMBU for metaprotemic analysis.

	PERIOD 1	***PERIOD 2***	***PERIOD 3***	***PERIOD 4***
COW 1	HPO	MAP	CTL	COS
COW 2	MAP	COS	HPO	CTL
COW 3	CTL	HPO	COS	MAP
COW 4	COS	CTL	MAP	HPO
GOAT 1	MAP	COS	HPO	CTL
GOAT 2	CTL	HPO	COS	MAP
GOAT 3	HPO	MAP	CTL	COS
GOAT 4	COS	CTL	MAP	HPO

Table 3.2. Chemical composition and formulation of concentrate and hay forage. Table obtained from Fougère et al. (2018) with modifications.

Item		**Concentrate**			**Forage**
	Control	COS	MAP	HPO	Grassland hay
Ingredient (g/kg of dry matter)					
Wheat		395			
Corn	532	394	518	500	
Soy	138	150	142	147	
Dehydrated alfalfa	275		283	294	
Molasses cane	37	35	38	39	
Dicalcium phosphate	2	2	2	2	
Carbonate flour	11	19	12	13	
Salt	3	3	3	3	
Mineral and vitamin complement	2	2	2	2	
Chemical composition (g/kg of dry matter)					
Organic matter	923	964	922	932	921
Crude protein	267	264	257	265	142
Neutral detergent fiber	198	125	206	206	625
Acid detergent fiber	110	42	113	116	351
Starch	365	507	342	337	
Ether extract	23	47	26	39	15
14:0	0,04	0,02	0,4	0,24	0,07
16:0	3,01	5,87	4,05	10,87	2,38
***cis*-9 18:1**	0,03	0,05	0,05	0,02	0,02
18:0	0,62	0,78	0,59	8,33	0,2
***cis*-9 18:1**	5,43	11,98	3,98	4,44	0,35
***cis*-11 18:1**	0,23	0,36	0,23	0,18	0,05
18:2n-6	12,6	24,8	8,7	9,7	1,98
18:3n-4	2,01	0,82	2,84	1,62	4,98
20:5n-3	0,003	0,001	0,002	0,002	nd*
22:5n-3	0,05	0,03	0,09	0,04	0,13
22:6n-3	nd	nd	3,06	nd	0,004
Total amount of fatty acids	26,3	45,8	27,1	37,3	12,98
Energy, MJ/kg of dry matter **	8,43	7,66	6,51	6,82	4,84
Protein, g of protein digestible in intestine/ kg of dry matter **	102	104	103	103	55

* – nd: Not detectable
** – Calculated according to INRA (2007)

Rumen fluid was sampled by stomach tubing from 8 animals, four cows and four goats, in the 4th week of every experimental 28-day period. All samples were collected 3-4 hours after feeding, snap-frozen in liquid nitrogen and stored at -80°C. In total, 32 rumen fluid samples (8 animals fed four diets) were sent to the Norwegian University of Life Sciences (NMBU) for metaproteomic analysis.

3.2 Cell lysis and protein extraction

The samples were prepared according to the bottom up proteomics approach for proteomic analysis, described in **Section 1.5.1.1** Sample preparation for Nano LC-MS/MS of the total 32 samples (2 animal groups x 4 diets x 4 biological replicates per diet) were done in three stages, with 10, 12 and 10 samples prepared in each stage, respectively. All three preparation stages were conducted in an equal manner, as described in the following sections.

3.2.1 Bead beating cell lysis and protein extraction

As the rumen microbiome consist of a variety of Gram-positive and Gram-negative anaerobic bacterial cells, archaeal, fungal and protozoal cells, bead beating, a relatively hash and mechanical cell lysis was used to disrupt the cells and extract proteins from the rumen fluid samples. Prior to lysis, the rumen fluid samples were thawed on ice until liquid and vortexed to homogenize the samples. One FastPrep® tube for every sample was marked with its respective sample number. Approximately 4 mm of glass beads (≤ 106 µm) were added to each FastPrep® tube. Furthermore, 150 µl of 3 x lysis buffer (See section 2.3 Buffers) with a high concentration of detergent (SDS), here 4%, was added to enhance denaturation of the proteins. Sodium dodecyl sulfate (SDS) is a negatively charged detergent that binds to the protein chain backbone, causing the protein to unfold. SDS also increases the solubility of proteins, typically yielding better protein recovery than lysis buffers without detergents. The method of adding a high concentrate detergent in the lysis buffer is based on (Zougman et al., 2014). 300 µl of each sample was added to the prepared Fast Prep tube, giving a total volume of 450 µl in each tube. FastPrep® tubes were briefly vortexed and let rest on ice for 30 minutes.

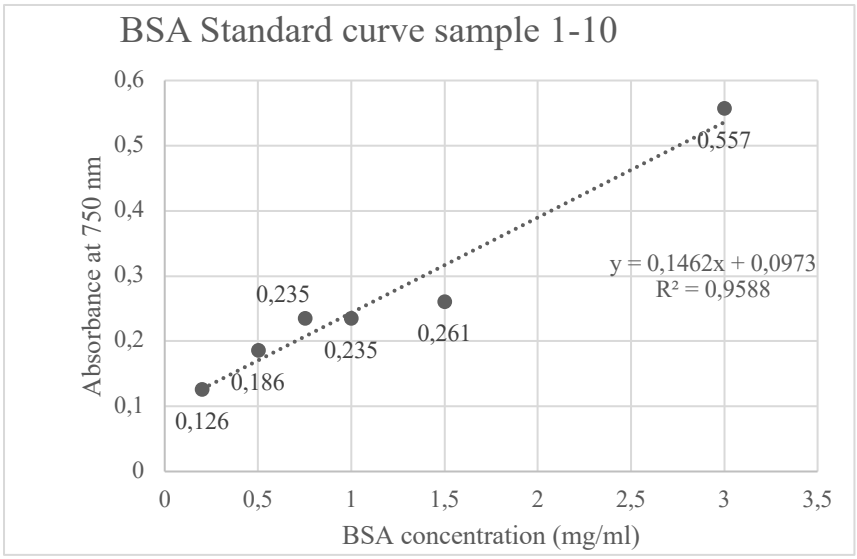
The cells were lysed through orbital shaking of the rubes containing beads and sample, causing a mechanical beating of cells and thereby releasing intracellular proteins and nucleic acids into solution (MP Biomedicals, 2018), using FastPrep-24™ Classic Grinder for 3 x 60 seconds at 4.0 meter/ second. As protein was released from the lysed cells, samples were centrifuged at 16 000 x *g* for 15 minutes at 4°C to pellet cell debris, leaving released proteins in the supernatant. Supernatant from every sample was carefully pipetted out and transferred to separate, marked Eppendorf Safe-Lock tubes. Sample lysates that were not handled immediately were frozen at -20° C until further preparation.

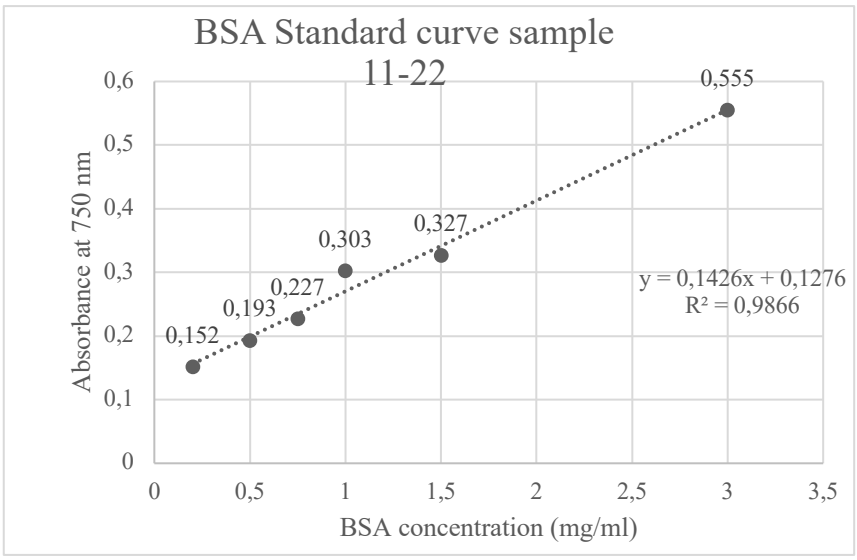
3.2.2 Measuring protein concentration

Protein concentration in sample lysates were measured using the Bio-Rad DC™ Protein Assay (Bio-Rad, 2019), which is detergent compatible (DC) and able to measure protein concentration in samples with 4% SDS. The protocol for microplate assay (Chapter 5.2) was followed as described in the Bio-Rad DC™ Protein Assay Instruction Manual. The color changing reaction in the Bio-Rad DC™ Protein Assay follows the Lowry assay principle first described in Lowry's highly cited paper from 1951, consisting of two steps (Bio-Rad, 2019; Lowry et al., 1951). In the first step, copper from the alkaline copper tartrate solution (Reagent A from the Bio-Rad DC™ Protein Assay Kit) reacts with peptide bonds to form a copper:protein complex. The reaction is followed by a reduction of Folin's phenol reagent (Reagent B from the Bio-Rad DC™ Protein Assay Kit) in the second step of the reaction. Folin's phenol is an oxidizing reagent which oxidizes the copper ion in the copper:protein complex, while at the same time being reduced, forming a blue color (Redmile-Gordon et al., 2013). This color absorbs light at 750 nm. The color change reaction reaches 90% of its maximum color development within 15 minutes and does not change more than 5% in one hour after addition of the reagents (Bio-Rad, 2019). A study conducted by Redmile-Gordon and colleagues in 2013 concluded that the Lowry method provides an accurate quantitative protein concentrate measurement, as it enables clear distinction between color development from protein and non-protein origin (Redmile-Gordon et al., 2013).

Bovine serum albumin (BSA) standards were made with concentrations of 0.2, 0.5, 0.75, 1.0, 1.5, 3.0 mg/ml BSA. Absorbance for all sample lysates, including BSA standards, was measured after 15 minutes at 750 nm (A_{750}) on BioTek™ Synergy™ H4 Hybrid Microplate Reader. A_{750} measured for the BSA standards were used to create a standard curve in Excel.

The regression equation was used to calculate the protein concentration in the sample lysates in Excel. As high protein concentrations were expected, the sample lysates were diluted in a ratio of 1:2 for samples 1-10 and 1:4 for samples 11-32. Standard curves for BSA standards are in **Figure 3.1** shown below.





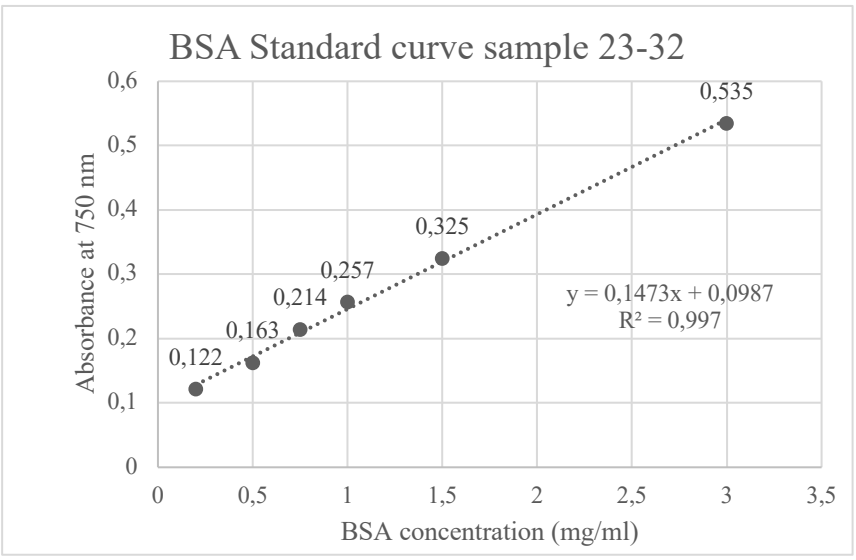


Figure 3.1. Standard curves for BSA standards for all 32 sample lysates with R^2 and regression equation used to calculate protein concentration in the sample lysates.

3.2.2.1 Troubleshooting: Protein concentration measurements

Initial attempts to measure protein concentration for the sample lysates were performed with the Bradford method for protein quantification, developed by M.M. Bradford in 1976 (Bradford, 1976). The Bradford protein assay relies on Coomassie Blue to bind to aromatic amino acids, such as phenylalanine, tryptophan and tyrosine (Kruger, 2009). The acidic assay reagent solution (Protein Assay Dye Reagent Concentrate, see **Section 2.2.2**), with a red color and an absorbance maximum at 450 nm, becomes more basic when Coomassie Blue binds to protein, causing a color change from red to blue and an absorbance shift from 450 nm to 595 nm (Kruger, 2009).

To 5 µl of lysate from each sample, 795 µl Tris-HCl (20 mM) and 200 µl Protein Assay Dye Reagent Concentrate were added in Axygen® 1.5 ml tubes. In addition to the rumen fluid samples, one blank sample (5 µl lysis buffer) and one control sample (5 µl BSA with known concentration) were made with the same ratio of reagents. A_{595} was measured on the Eppendorf BioPhotometer D30 after five minutes. Notably, high concentrations of detergent, such as SDS in the lysis buffer, interferes with the Protein Assay Dye Reagent Concentrate used for Bradford protein assay by binding to protein and thereby competing with Coomassie Blue (Kruger, 2009). This causes a underestimation of protein concentration in the sample (Kruger, 2009). The Bio-Rad DC™ Protein Assay is detergent compatible (DC), i.e. does not interfere with SDS (Bio-Rad, 2019), providing a more accurate estimate of protein content in each sample. Thus, the final protein concentration for the sample lysates was measured using the Bio-Rad DC™ Protein Assay.

3.3 Sodium dodecyl-sulfate polyacrylamide gel electrophoresis

The proteins were cleaned up by sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The negatively charged detergent, SDS, binds to the protein backbone (Medical & Biological Laboratories CO., 2017), while a strong reducing agent, i.e. dithiolthreiol (DTT), is used to reduce disulfide bonds formed in cysteine in proteins prior to electrophoresis (NCBI, 2019). The binding of SDS and addition of reducing agent (DTT) causes the protein to unfold into linear chains with negative charge, enabling proteins to be separated based solely on their chain length (Medical & Biological Laboratories CO., 2017). The negative charged SDS-

bound, reduced protein migrate towards a positive anode. Smaller proteins migrate easier through the mesh-like polyacrylamide gel, while larger proteins migrate shorter because of resistance in the gel.

3.3.1 Protein clean-up by SDS-PAGE

40 µg of protein from every sample between sample 1-10 and 50 µg of protein from every sample between sample 11-32 were loaded on the SDS-PAGE gels. The regression equation from standard curves from BSA standards was used to calculate the protein concentration in each sample, and to load approximately the same amount of starting material for each sample. 40 µg has previously been shown to be sufficient for MS. By dividing the amount of required protein on the gel (µg) by the protein concentration measured (µg/µl), the amount of each sample to add to the gel (µl) was calculated. The calculated amount of lysate from every samples was transferred to a new, marked Axygen® 1.5 ml tube, together with 3 µl of NuPAGE™ Sample Reducing Agent (10X) and 7.5 µl of Novex™ NuPAGE™ LDS Sample Buffer (4X). NuPAGE™ Sample Reducing Agent contains 500 mM dithiothreitol (DTT) (Thermo Fischer Scientific, 2019) and reduces disulfide bonds in proteins, as mentioned above. The maximum volume possible to add in each well on the Mini-PROTEAN® TGX Stain-Free™ Gel is 30 µl. Given these fixed amounts of the other reagents, the sample lysate volume could not exceed 19.5 µl. It was therefore desired to have as high protein content in each sample lysate as possible to ensure reaching 40 µg within the 19.5 µl.

Milli-Q was added to reach a total volume 30 µl in the cases where less than 19.5 µl of sample was used. Samples were heated in a water bath for 5 minutes at 99°C. The gels were placed in the Mini-PROTEAN® Tetra Cell system. The electrode gasket was filled with fresh 1 x TGS-buffer (Tris/Glycine/SDS Buffer, Bio-Rad) and the rest of tank was filled with used 1 x TGS-buffer. 28 µl of each prepared sample was carefully loaded on Mini-PROTEAN® TGX Stain-Free™ gels and caution was used not to spill between wells. In addition to a BenchMark™ Protein Ladder, 5 samples were loaded per 10 well gel with blanks in between to inhibit one sample well to contaminate the neighboring well, as the volume added into each well were close to the maximum volume for this type of gel. The gel was run in the Mini-PROTEAN® Tetra Cell systems for about 2-3 minutes on 270 V and 400 milliampere (mA) with the PowerPac™ Basic Power Supply, until the whole sample had migrated about 1 cm down into the gel. For

this thesis, SDS-PAGE was not used as a protein separation step, but rather as a clean-up step to get rid of e.g. humic substances and interfering compounds which remain in the gel.

3.3.2 Staining and destaining of gels

The gel was further removed from the gel chamber, carefully transferred to a gel staining box and stained with stain solution (See **Section 2.3**) for 1 hour on slow shaking (30 times/ minute) on the IKA® HS 260 Basic shaker. The stain solution contains Coomassie Brilliant Blue R-250, which binds to proteins and make proteins visible in the gel. After 1 hour the stain solution was gently removed, and the gel was destained in a destain solution (See **Section 2.3**) for 20 minutes and placed back on slow shaking. The destain solution was gently poured of and the gel was destained again in destain solution for additional 20 minutes on shaker. If necessary (in case the gel was still noticeably stained/ blue), the destain process was repeated once more, for a maximum time of 1 hour. When desired result (transparent/pale blue gel with clearly visible blue bands) was obtained, the destain solution was replaced with 50/50 Milli-Q/destain solution and placed on slow shaking overnight. The following day the liquid was gently poured of and replaced with Milli-Q.

After staining of the gels, the visible band from each sample was carefully excised from the gel and further divided into 1x1 mm pieces with a scalpel and transferred to marked Eppendorf Safe-Lock tubes. Within this process, was crucial to prevent contamination from one sample to another. The scalpel was washed with Milli-Q between each excision and the gel pieces from one sample was carefully gathered in a marked Eppendorf Safe-Lock tube prior to cutting out another sample.

3.3.3 De-coloring and cleaning of gel pieces

200 µl Milli-Q was added to each Eppendorf Safe-Lock tube with gel pieces and the tubes were incubated on ThermoMixer® on 22°C on shaking (800 rpm) for 15 minutes. The liquid was subsequently removed and 200 µl 50% acetonitrile (ACN) / 25 mm ammonium bicarbonate (AmBic) was added to each tube containing gel pieces and incubated on ThermoMixer® on 22°C on shaking for another 15 minutes. This step was repeated once. After removing 50% ACN / 25 mm AmBic, 100 µl 100% ACN was added to each sample. The samples were further incubated on ThermoMixer® on 22°C on shaking for 5 minutes. The liquid was removed, and the white and shrunken gel pieces were left to air dry for 1-2 minutes.

3.3.4 Reduction and alkylation

50 µl DTT solution (See section 2.3) was added to each sample and the samples were incubated on ThermoMixer® on 56°C on shaking for 30 minutes. Dithiothreitol (DTT) was used to reduce disulfide bonds in cysteines in the proteins. The samples were cooled down and the excess liquid collected by spinning in Heraeus™ Pico™ 21 centrifuge at room temperature for 30-60 second on 14 000 x *g*, after which the liquid was removed. 50 µl IAA solution (See **Section 2.3**) was added to each sample and the samples were further incubated in the dark for 30 minutes at room temperature. Iodoacetamide (IAA) is an alkylating agent that binds to the thiol group of cysteines, thereby preventing cysteines to form disulfide bonds (Sigma, 2001). IAA is sensitive to light and the IAA solution was therefore made fresh for each of the three sample preparations. After incubation, IAA was removed and 200 µl 100% ACN was added to the samples. The samples were further incubated for on ThermoMixer® on shaking on 22°C. The liquid was removed, and the white and shrunken gel pieces were left to air dry for 1-2 minutes.

3.4 In gel-digestion

After the above reduction and alkylation of the proteins, the proteins were digested into peptides by trypsin. Trypsin is a specific and non-selective serine protease, which cleaves the carboxyl end of lysine and arginine (unless lysine or arginine is followed by a proline) in the protein chain. 30 µl of 10 ng/µl trypsin solution (See **Section 2.3**) was added to each sample. The samples were incubated on ice for 30 minutes. If necessary, more trypsin buffer (See **Section 2.3**) was added to completely cover the gel pieces. The samples were incubated overnight on ThermoMixer® on shaking on 37°C. The follow day, the samples were cooled down and added 1% trifluoroacetic acid (TFA) to stop the trypsin digestion. Further, the samples were sonicated on Branson 3510 Ultrasonic Cleaner for 15 minutes to release peptides from the gel pieces into the TFA solution.

3.5 Peptide clean-up using ZipTips and centrifugation

The peptides were concentrated and eluted using ZipTips® Pipette Tips. ZipTips have a C18 hydrophobic stationary phase for easy purification and concentration of low concentrations of peptides for sensitive data analysis methods, such as mass spectrometry analysis (Merck Millipore, 2019). ZipTips were placed in pipette tip adapters in Eppendorf Safe-Lock tubes. The tubes were marked with sample numbers, as one tip was only used for one sample, and

placed in the Heraeus™ Pico™ 21 centrifuge. The ZipTips were conditioned with 10 µl MeOH and the liquid was spun through the column for 5 seconds at 100 x *g*. The ZipTips were further conditioned with 70% ANC/ 0.1% TFA and the liquid was spun again for 5 seconds at 100 x *g*. The ZipTips were equilibrated with 10 µl 0.1% TFA and spun for 5 seconds at 100 x *g* to ensure that all liquid had passed through the column.

50 µl of each sample was added to the its respective ZipTip in the marked tube corresponding to the sample number. The sample were spun down for 15 seconds at 500 g. The peptides, now bound to the column, were washed with 10 µl 0.1% TFA. The ZipTips were spun at 100 g for another 5 seconds. The Eppendorf Safe-Lock tubes with conditioning and washing liquids were discarded to waste and ZipTips with bound peptides were placed in marked tubes in the Heraeus™ Pico™ 21 centrifuge for elution of the peptides. The peptides were eluted with 20 µl 70% ACN/ 0.1% TFA. The liquid was spun down in the centrifuge at 100 x *g* for 10 seconds. The ZipTips were checked between each washing step to ensure that liquid had passed all the way through the column and out into the tube. The liquid from the elution step was carefully pipetted out and transferred to marked Snap Ring Micro-vial HPLC vials.

3.6 LC-MS/MS analysis

All 32 HPLC vials with samples were dried in a centrifugal evaporator **(**vacuum concentrator) for 15 minutes or until completely dry (no visible liquid) at 45 °C. Dried peptides were dissolved in 10 µl 2% ACN/ 0.1% TFA and half of this was injected on the UltiMate™ 3000 RSLC nano liquid chromatography mass spectrometer (Nano LC-MS/MS). Chromatographic separation was done using a Acclaim™ PepMap™ 100 column (C18, 3 µm, 100 Ångstrøm (Å)) with 50 cm bed length. The gradient for HPLC was two hours. The flow rate was 300 nL/min and the solvent gradient was 3.2% Solvent B (96.8% Solvent A) in two minutes to 9.6% Solvent B. The Solvent B gradient increased to 34.4% in 93 minutes, and further increased to 48% in four minutes and then directly to 80% Solvent B for five minutes, before Solvent B was decreased to 3.2% for equilibration. The peptides were analyzed by nano-LC-MS/MS using a Q-Exactive hybrid quadrapole Orbitrap MS, as described in **Section 1.5.1.1**.

3.6.1 Optimalization of rumen associated databases

Because a sample specific protein sequence database was not available for the sample data provided by INRA at the time of the thesis, three databases were constructed. MS/MS raw data was search against these three different databases, hereby referred to as RUmen DataBase (RUDB) 1, 2, and 3 and shown in **Table 3.3**. The databases were with the attempt to generate a pseudo-metagenomics database, i.e. trim a larger non-specific database to best reflect the organisms present in the samples, and thereby be able to capture changes in microbial composition in the rumen microbiome and their metabolic functions when the ruminant is fed different types of feed (i.e. identified in multiple replicates). A similar approach has been suggested by the metaproteomics community in the lack of a true sample-specific database in previous studies (Muth et al., 2016; Tanca et al., 2016). The first database, called RUDB1, consisted of 425 559 protein entries. It included Metagenome-Assembled Genomes (MAGs) from two previous rumen genome studies, respectively Hess et al. (2011) and Parks et al. (2017). In addition, RUDB1 included viral peptides, fungal genomes and 13 cultivated genomes from the rumen genome database Hungate 1000 (Seshadri et al., 2018) in addition to the genome of *Fibrobacter succinogenes* subsp. *Succinogenes* S85 (NCBI accession: NC_013410) and the methanogenic *Methanobrevibacter ruminantium* strain M1 (NCBI accession: NC_013790.1). The reduced database, called RUDB2, consisted of 245 422 protein entries, of which the viral peptides and fungal genomes were excluded. In addition to the MAGs and cultivated bacterial genomes from RUDB1, 4 methanogenic archaeal genomes from the Hungate1000 database was added to RUDB2. For RUDB3, we took advantage of a recently developed software pipeline in Galaxy (not yet published) that was able to take an initial large database and comprehensively trimming this using iterative sectioning and MS/MS searching. This Galaxy sectioning workflow, developed by our collaborators, Tim Griffin and Praveen Kumar from the University of Minnesota, was used to create RUDB3. Briefly, the Galaxy sectioning workflow takes in a large amount of protein entries, in this case 1 293 596 protein entries, comprised of MAGs from the two mentioned studies, the five fungal genomes, 311 cultivated genomes from the Hungate1000 database and the genome of *F. succinogenes.* Moreover, RUDB3 included *Methanobrevibacter ruminantium* and 15 additional methanogenic archaeal genomes, both complete genomes and whole genome shotgun sequences from Li et al. (2019) and Henderson et al. (2015). In the Galaxy sectioning workflow, the protein entries are split into several sections that are individually searched and evaluated using SearchGUI and PeptideShaker. The identified proteins are being used to build a final

database while additional noise/ protein entries without any matching MS/MS data are added to avoid overfitting of the database to the data. This prevents bias in the final results. After sectioning, the final RUDB3 consisted of 303 834 protein entries,

Table 3.3. Content of the rumen databases created for protein identification.

RUDB1	**RUDB2**	**RUDB3**
101 MAGs	101 MAGSs	101 MAGs
13 Hungate genomes + *F. succinogens*	13 Hungate genomes + *F. succinogens*	311 Hungate genomes* + *F. succinogens*
1 added methanogen	4 added methanogens	16 added methanogens
5 Fungal genomes		5 Fungal genomes
913Viral scaffolds		
In total:		
425 559 protein entries	235 422 protein entries	303 834 protein entries

*Four of these 311 genomes from the Hungate database are methanogens, resulting in a total of 20 genomes from non-redundant methanogens in RUDB3.

3.6.2 Protein quantification and identification with proteomics software

Raw data from MS was search against the RUDBs using MaxQuant version 1.6.3.3 and the detected protein groups were subsequently explored in Perseus version 1.6.1.1. The RUDBs were supplemented with contaminant protein entries, such as human keratin, trypsin, and bovine serum albumin and concatenated with reversed sequences of all protein entries to enable estimation of the FDR. Hits to any of the contaminants, reversed sequences and protein entries that were identified only by site were removed. Oxidation of methionine, protein N-terminal acetylation, deamination of asparagine and glutamine and conversion of glutamine to pyroglutamic acid were used as variable modifications. Carbomidomethylation of cysteine residues were used as fixed modification. Trypsin were chosen as the digestive enzyme and maximum one missed cleavage was allowed. Additionally, more relaxed FDRs (5% and 10%) were applied to see the effect on the final results for one of the RUDBs. The first MaxQuant run (MQrun1) was carried out with RUDB1 with FDR 1%. The next three MaxQuant runs (MQrun2-4) were carried out with the RUDB2, with respectively FDR 1%, 5% and 10%. The

last MaxQuant run (MQrun5) was carried out with RUDB3 with FDR 1%. The five MaxQuant runs will hereafter be referred as MQrun1-5. For all runs, biological replicates of the 32 samples were grouped into a total of 8 groups based on their diet; cows fed CTL, HPO, MAP and COS, as well as goats fed CTL, HPO, MAP and COS. For a protein to be considered valid, it was required that the protein was identified and quantified in at least 2 of the 4 biological replicates. Functional annotation of the protein entries was done using InterProScan 5 version 5.32-71.0 (Fraser, 2017), and assigned to the detected protein groups in Perseus. All heatmaps with hierarchical clustering presented below were made with Euclidean distance (default), 30 clusters and 10 iterations for both row trees and column trees, with average linkage, no constraints and preprocessed with k-means, i.e. default settings.

3.6.3 Evaluation of host (Bos taurus) contamination

In addition to the five MaxQuant runs already described, there was conducted a sixth MaxQuant run, including only the raw files from samples from cow to check for host contamination. The run was set up as equal to MQrun5. In addition to RUDB3, the complete genome of cow (*Bos taurus*) was included. The complete genome of cow consisted of 37 312 protein entries and was downloaded from the UniProt website (UniProt proteome ID: UP000009136). The complete database (RUDB3 + complete cow genome) consisted of total 341146 protein entries.

4 Results

Metadata measurements were conducted for all samples by INRA. Animals fed the COS diet, basal concentrate supplemented with corn oil and wheat starch, had the lowest average methane production measured for both cow and goat (**Table 4.1**). Cows fed the COS diet had the highest percentage of propionate production. Standard deviations for goats fed CTL and COS were higher than the rest.

Table 4.1. The four diets fed to four cows and four goats in four replicates (32 samples) and their associated metadata. Here shown total VFA sum, in addition to sum of respectively acetate, propionate and butyrate and methane averaged for biological replicates for cow and goat fed the four diets. Standard deviation for CH_4 measurements was also included. All metadata measurements were conducted by INRA. Diet acronyms are as follows; CTL = basal control, no additional oil, HPO = basal concentrate supplemented with hydrogenated palm oil, MAP = basal concentrate supplemented with marine algae powder, COS = basal concentrate containing wheat starch and supplemented with corn oil.

		VFA sum mmol/l	**Acetate (% sum)**	**Propionate (% sum)**	**Burytate (% sum)**	**CH4 g/kg dry matter intake**	**Standard deviation CH_4 g/ kg**
COW	CTL	61,83	72,02	14,21	10,80	19,51	2,185
	HPO	70,89	72,52	14,71	9,86	16,90	2,561
	MAP	84,05	72,22	14,95	9,97	17,99	2,423
	COS	68,055	69,115	21,315	6,393	14,448	2,892
GOAT	CTL	33,70	65,73	15,62	13,66	19,52	3,958
	HPO	37,92	65,40	16,00	13,21	18,96	2,123
	MAP	59,46	62,97	20,38	12,08	18,31	2,595
	COS	25,27	64,17	19,07	9,90	13,47	4,144

A total of 599 593 MS/MS spectra was submitted for analysis of MS/MS raw data in MaxQuant. The number of MS/MS submitted spectra was the same for all of the MaxQuant database searches, except the sixth run, which are described in **Section 3.6.1** and in **Table 3.3**.

4.1 Database searches in MaxQuant

All five MaxQuant runs (MQrun1-5) were conducted and set up in an equal manner, as described in **Section 3.6.1**. From MQrun1, 33 637 MS/MS spectra, of the submitted 599 593 MS/MS spectra, were matched to a peptide sequence (**Table 4.2**). The identification rate

reflects the fraction of MS/MS spectra that were matched to a peptide sequence, providing an MS/MS identification rate of 5.61% for MQrun1, amounting to 1420 protein groups being identified. After removing hits identified as contaminants, hits identified from the reverse database and protein groups identified only by site, the number of identified protein groups were reduced to 1251. It was assumed that animal groups and individual animals had no effect on change in microbial composition, i.e. change in microbial composition were solely due to effect of the different diets, hence protein quantities (LFQ intensities) from biological replicates were averaged in Perseus, resulting in 8 groups (2 animal groups x 4 diets). As aforementioned, for a protein to be considered valid, it was required that a protein group was identified in at least 2 of 4 replicates within each group. This yielded in total 560 protein groups using an FDR of 1% with RUDB1 (**Table 4.2**). MQrun2 yielded 33 083 MS/MS spectra and an MS/MS identified rate of 5.52%, slightly less than MQrun1. Despite this, a total of 1577 protein groups were identified (**Table 4.2**). After filtration as above, the number of identified protein groups was reduced to 1265. The final filtering on biological replicates gave a total of 545 protein groups. By increasing the FDR to 5% in MQrun3, the number of identified MS/MS spectra increased to 36 755, yielding a higher MS/MS identified rate (6.13%) and a total of 2287 protein groups (**Table 4.2**). After initial and final filtration, the number of identified protein groups was reduced to 2082 and 579, respectively. MQrun4 was performed with a further increased FDR (10%) and identified 38 804 MS/MS spectra, yielding an MS/MS identified rate of 6.47%. 2511 protein groups were initially identified, reduced to 2169 after initial filtration, and to 613 protein groups after final filtration. In the final run, MQrun5, the MS/MS spectra were search against the sectioned database (RUDB3), 43 435 MS/MS spectra were identified, resulting in an MS/MS identified rate of 7.24%. In total, 2445 protein groups were identified. After removing hits identified as contaminants, hits identified from the reverse database and proteins identified only by site, the number of identified protein groups was reduced to 2211. Requiring that a valid protein was identified in at least 2 of 4 replicates, gave a total of 838 identified protein groups (**Table 4.2**).

Table 4.2 Overview of MaxQuant runs with RUDBs. The database, database size and the FDR used, the number of identified MS/MS spectra, the rate of identified MS/MS spectra and the number of identified and valid protein groups identified in at least 2 of 4 replicates for all five MaxQuant runs with different RUDBs.

Run	**MQrun1**	**MQrun2**	**MQrun3**	**MQrun4**	**MQrun5**
Database	RUDB1	RUDB2	RUDB2	RUDB2	RUDB3
Protein entries	425559	235422	235422	235422	303834
FDR	1%	1%	5%	10%	1%
MS/MS Identified	33647	33083	36755	38804	43435
MS/MS Identified rate	5.61%	5.52%	6.13%	6.47%	7.24%
Protein groups, prior filtration	1420	1577	2287	2511	2445
Protein groups, after initial filtration	1251	1265	2082	2169	2211
Protein groups, after final filtration*	560	545	579	613	838

* requires a protein group to be identified in at least 2 of 4 replicates within each group

As shown in **Table 4.2**, MQrun2 yielded the lowest MS/MS identified rate and the least protein groups. Even though RUDB1 is substantially larger in size than RUDB2, MQrun1 identified more MS/MS spectra and more protein groups than MQrun2. The amount of MS/MS spectra and protein groups identified increased in MQrun3-4, as the FDR increased. MQrun5 identified the largest amount of MS/MS spectra and hence had the highest MS/MS identified rate. The most identified valid protein groups, after filtration, was also identified in MQrun5. Because RUDB3 and MQrun5 identified the most protein groups of the three databases tested, the downstream analysis of the protein groups in this thesis was conducted based on this run.

4.2 Protein quantification

After the abovementioned evaluation of each MaxQuant run, the resulting protein groups from MQrun5 was quantified and statistically analyzed in the Perseus matrix. The abundance values for each of the 838 identified protein groups are presented as log2-transformed LFQ intensities, representing the relative protein quantifications, as described in **Section 1.5.1.2**. The LFQ intensities from each sample are presented in a histogram in **Figure 4.1** and shows that samples originating from goat had lower protein quantities or proteins not identified at all, compared to samples originating from cow.

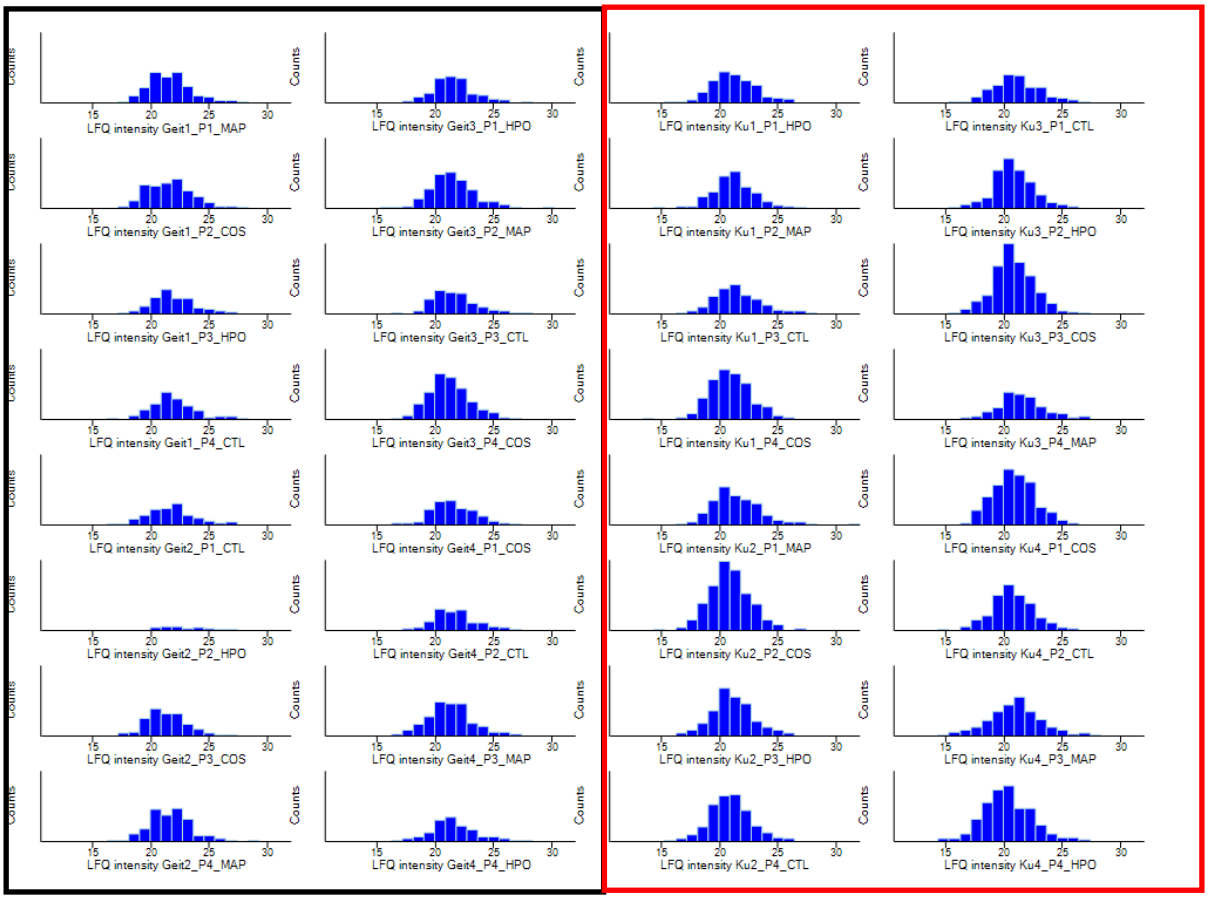


Figure 4.1. Histograms of identified protein quantifications from MQrun5 (log 2 transformed-LFQ intensity) for all 32 samples. LFQ intensities from each sample are named with the diet (CTL/ HPO/ MAP/ COS) fed to what animal (cow: Ku 1-4, Goat: Geit 1-4) and in which period (P1-P4) the diet was fed to the animal. Samples originating from goat (marked in black box) had fewer identified proteins and more missing hits compared to samples originating from cow (marked in red box). X-axis represent distribution of protein groups with low to higher LFQ intensities. Y-axis represent the number of identified protein groups with given LFQ intensity. Diet acronyms are as follows; CTL = basal control, no additional oil, HPO = basal concentrate supplemented with hydrogenated palm oil, MAP = basal concentrate supplemented with marine algae powder, COS = basal concentrate containing wheat starch and supplemented with corn oil.

This was also observed when visualizing the averaged LFQ intensities of the 8 groups of biological replicates (2 animal groups x 4 diets) for the 838 protein groups identified in MQrun5 in a heatmap (**Figure 4.2)** below. Here, the color grey indicated non-identified protein groups for one biological replicate. The protein profile from cow and goat formed two distinct clusters, suggesting that the relative protein quantification patterns are different between cow and goat, and might also reflect a different ruminal microbiome.

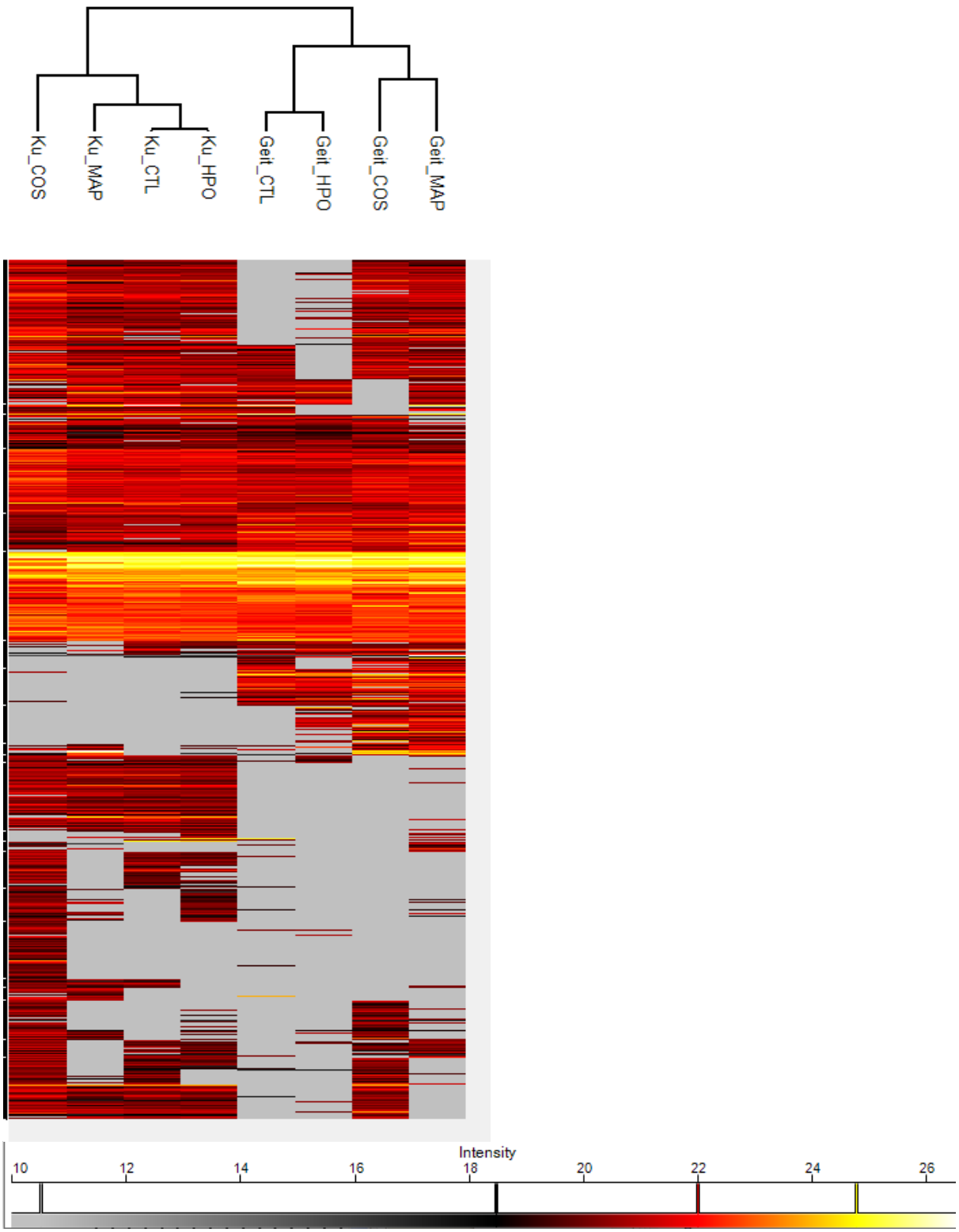


Figure 4.2 Averaged log 2-transformed LFQ intensities across four biological replicates in both animal group for 838 protein groups identified from MQrun5. Grey indicates missing hits, while black-to-red indicates low to medium high protein quantity and yellow-to-white indicates high quantity protein groups. Color bar is shown in the bottom of the figure. Biological replicates are grouped in 8 groups, where cows fed COS are called Ku_COS, cows fed MAP are Ku_MAP, cows fed CTL are Ku_CTL, cows fed HPO are Ku_HPO, goats fed CTL are Geit_CTL, goats fed HPO are Geit_HPO, goats fed COS are Geit_COS, and goats fed MAP are Geit_MAP. Samples originating from goat had fewer identified protein groups compared to samples originating from cow. Diet acronyms are as follows; CTL = basal control, no additional oil, HPO = basal concentrate supplemented with

hydrogenated palm oil, MAP = basal concentrate supplemented with marine algae powder, COS = basal concentrate containing wheat starch and supplemented with corn oil.

As both **Figure 4.1** and **Figure 4.2** showed, visibly fewer identified protein groups and more missing hits from samples originating from goat. Further work, such as downstream analysis of protein quantification and functional context in this thesis was conducted using only results from samples from cow.

Scatter plot with Pearson correlations for biological replicates for samples from only cows fed the low-methane COS diet are shown in **Figure 4.3**, where the lowest correlation was found between Cow 1 and Cow 4 fed the COS diet, and the highest correlation was found between Cow 1 and Cow 3 fed the COS diet. Scatter plot with Pearson correlation for biological replicates for samples from only cow fed the all four diets are shown in **Appendix A-1**.

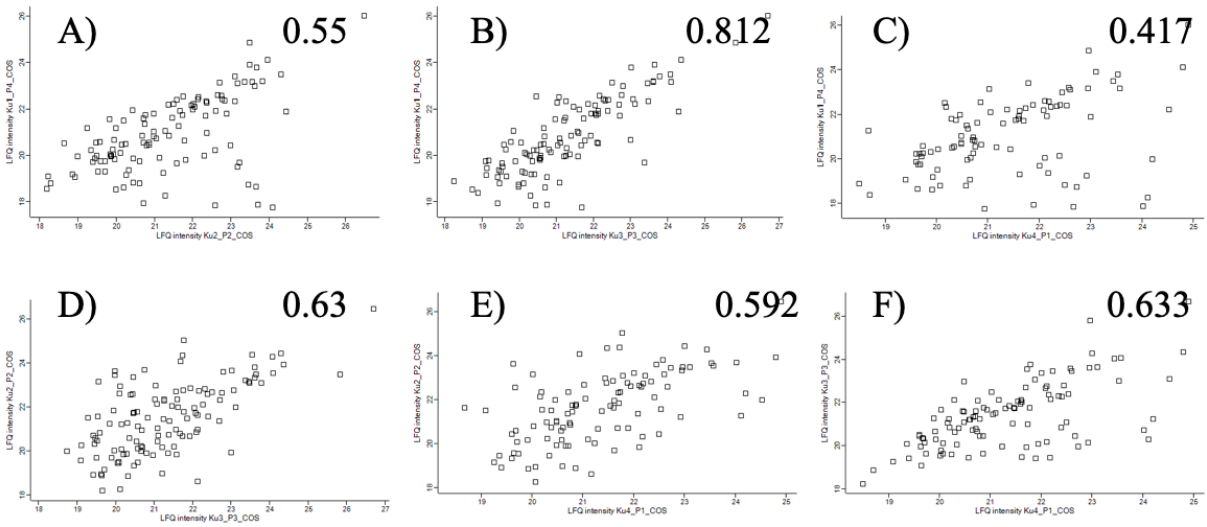


Figure 4.3. Scatterplots with Pearson correlations for biological replicates for samples where the cow was feed the low-methane COS diet identified in MQrun5. A) Scatterplot with Pearson correlation between Cow 1 and Cow 2 fed the COS diet, at 0.55. B) Scatterplot with Pearson correlation between Cow 1 and Cow 3 fed the COS diet, at 0,812. C) Scatterplot with Pearson correlation between Cow 1 and Cow 4 fed the COS diet, at 0,417. D) Scatterplot with Pearson correlation between Cow 2 and Cow 3 fed the COS diet, at 0.63. E) Scatterplot with Pearson correlation between Cow 2 and Cow 4 fed the COS diet, at 0.592. F) Scatterplot with Pearson correlation between Cow 3 and Cow 4 fed the COS diet, at 0.633.

The sixth MaxQuant run (MQrun6) mentioned in **Section 3.6.1** included the complete genome from cow (*Bos taurus*) to check for the presence of host proteins in the samples. Because further analysis was conducted only using the samples from cow, host contamination from goat (*Capra aegagrus hircus*) was not conducted. Because only the samples from cow were included in this

run, only MS/MS spectra submitted from samples from cows, were used in this analysis. 275226 MS/MS spectra were submitted for analysis. 29661 MS/MS spectra were identified, yielding a MS/MS identified rate of 10.8%. After removing hits identified as contaminants, hits identified from the reversed database and protein groups identified only by site, this run yielded 2125 identified protein groups, where 198 of them originated from the host. Requiring that a valid protein was identified in at least 2 of 4 replicates (16 samples, i.e. four cows fed each of the four diets), gave a total of 812 identified protein groups, where 70 protein groups originated from the host.

4.3 Taxonomic and functional annotation

Taxonomic and predicted gene function information from the MAGs and isolate genomes that were used to create peptide sequence databases was also used to manually annotate all of the identified protein groups from MQrun5 in Excel. As previously mentioned, functional annotation of the database was done with InterProScan 5 version 5.32-71.0, and further assigned to the protein groups in Perseus. Even though RUDB3 was based on genomic information, phylogenetic annotation was only available for 716 838 valid protein groups identified in MQrun5. The genomic information was downloaded in 2017 and updated genomes since then were therefore not phylogenetically annotated. 396 of the 716 protein groups, that taxonomical annotation was available for, were assigned to a gene ontology (GO) term, making up 72 different groups of biological processes (**Figure 4.4**). The most abundant GO term was oxidation-reduction processes, assigned to 99 protein groups.

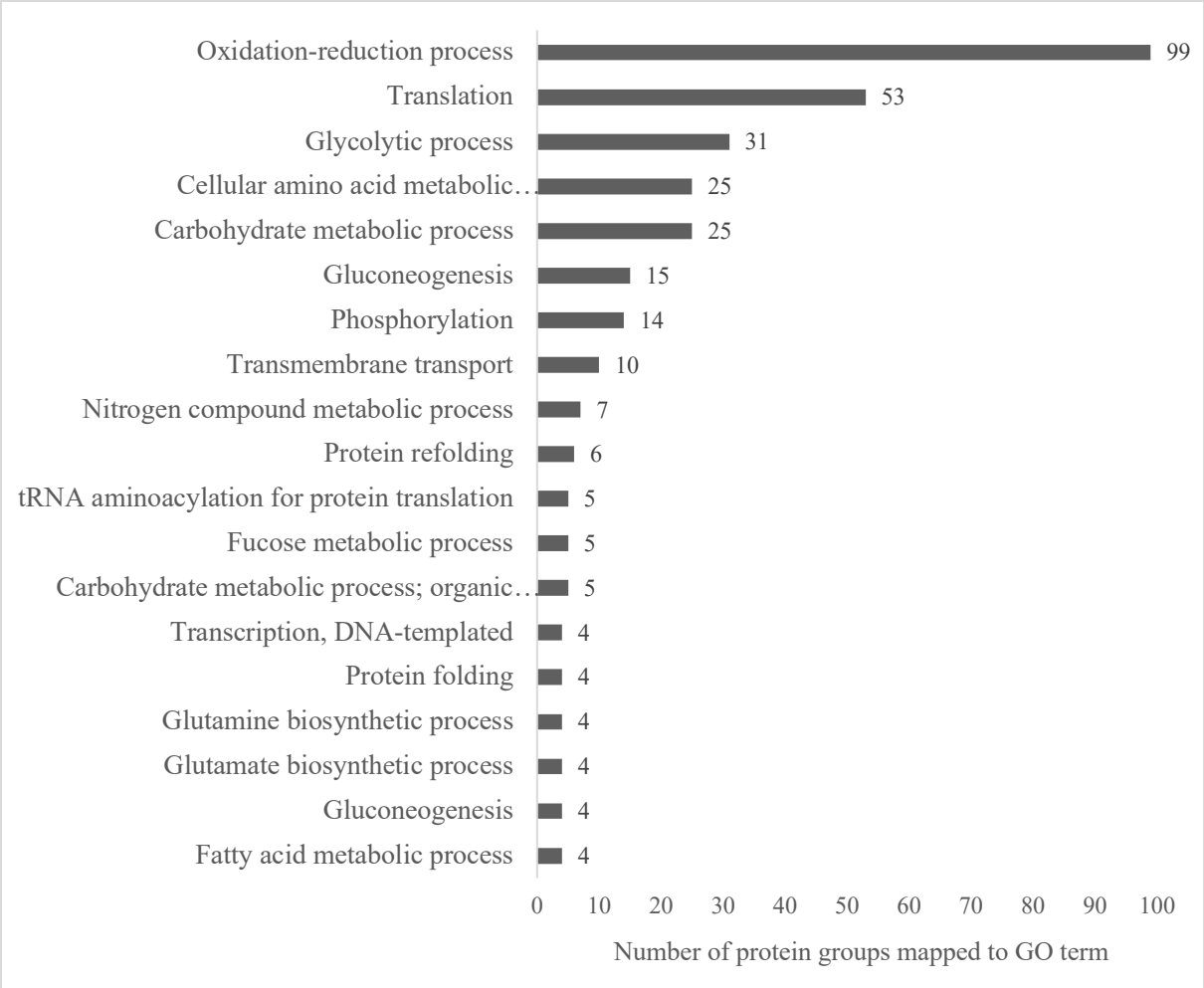


Figure 4.4. Distribution of gene ontology (GO) terms assigned to identified protein groups with taxonomic annotation from MQrun5. In total 72 different GO terms could be mapped for in total 396 protein groups assigned to a GO term. GO terms categories assigned to less than 4 protein groups are not shown in this figure. This figure shows the 19 most abundant GO terms amongst the 72 GO terms that protein groups were assigned to. These 19 GO groups were assigned to 324 of the 396 protein groups that GO terms were available for.

Of the 716 protein groups that taxonomic annotation was available for, 691 protein groups were assigned to bacterial taxa, 3 protein groups originated from archaeal taxa and 22 protein group were assigned to fungal taxa (**Figure 4.5A**). The 691 protein groups assigned to bacteria, were distributed amongst 7 taxa, where most of the protein groups originated from *Bacteroidetes* and *Firmicutes*, as shown in **Figure 4.5B**. The 22 fungal protein groups identified were distributed amongst four of the five fungal genomes included in the database (*Anaeromyces*, *Neocallimastix*, *Orpinomyces*, and *Piromyces*) of which *Piromyces* demonstrated the highest number of detected protein groups (8 protein groups) (**Figure 4.5C**)

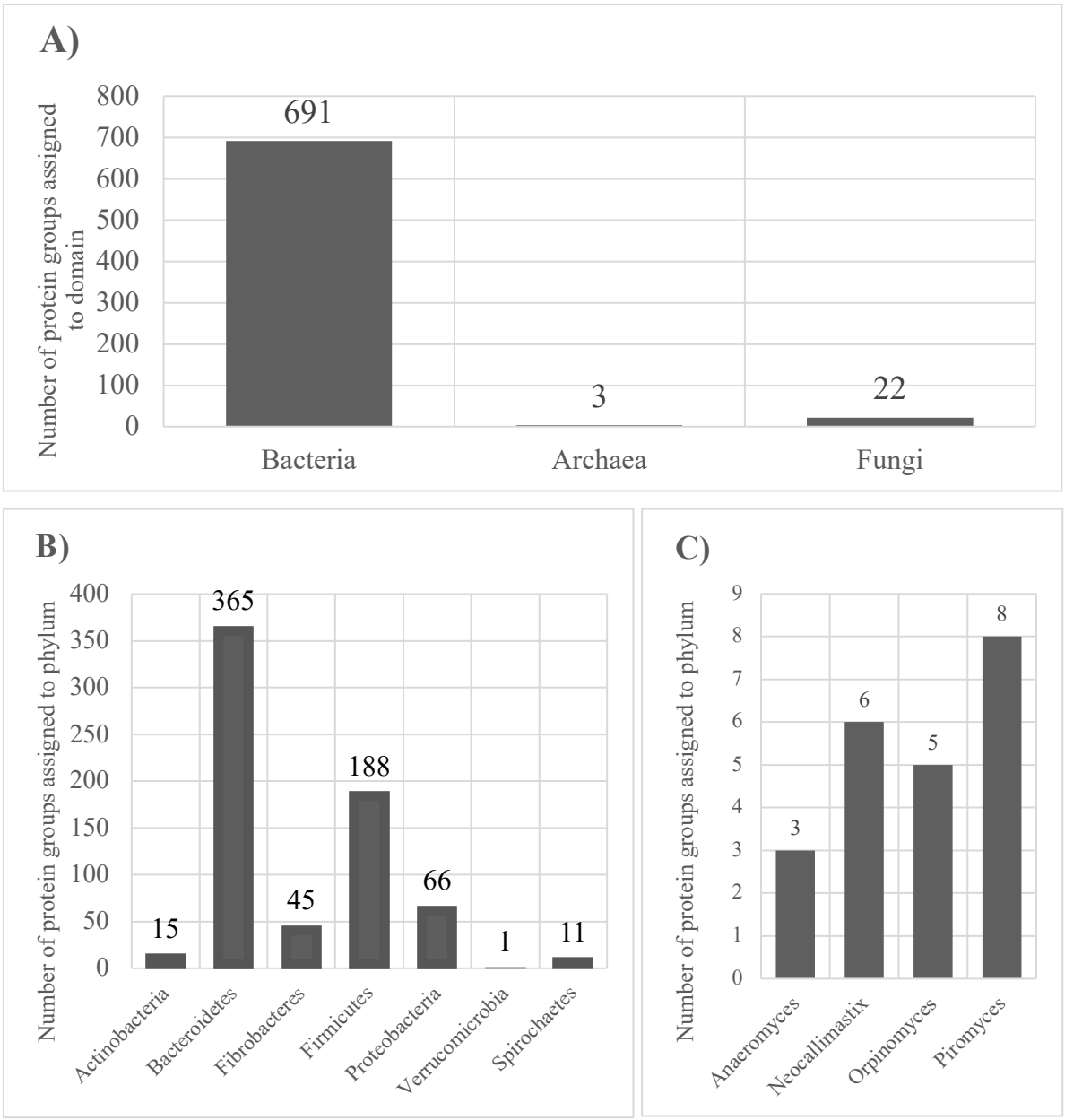
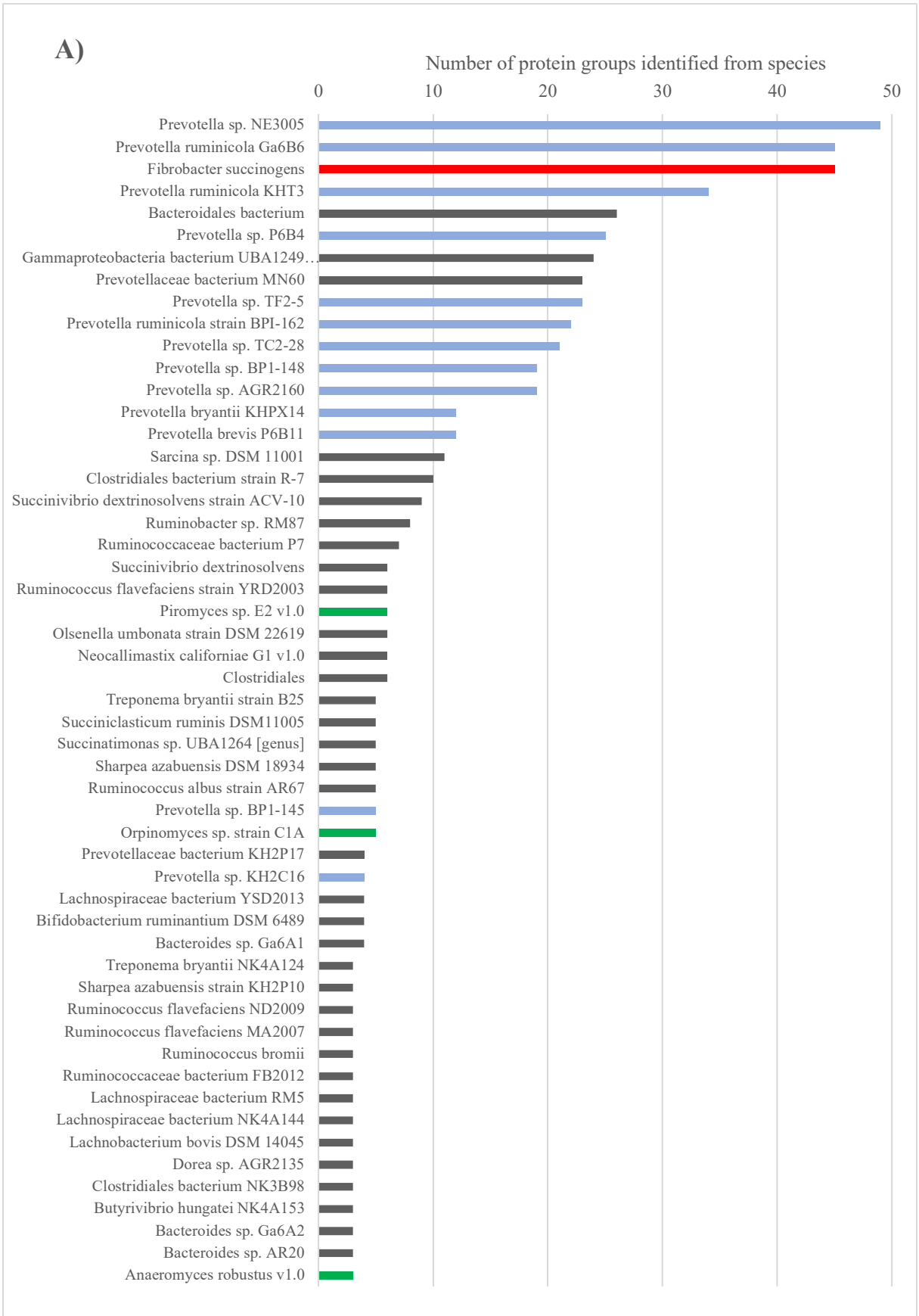


Figure 4.4 Phylogenetic distribution of the protein groups. A) The protein groups detected in MQrun5 assigned to different biological domain. Only three protein groups were assigned to the methanogens *Methanobrevibacter* and *Methanosarcina*. B) The distribution of 691 protein groups assigned to 7 bacterial taxa. C) The distribution of 22 protein groups assigned to four fungal taxa.

Although the genomes of 20 methanogens were included in RUDB3, only three protein groups were assigned to archaeal taxa derived from the phylum *Euryarchaeota,* and these were taxonomically annotated to the methanogenic genera *Methanobrevibacter* and *Methanosarcina*. The 716 protein groups that taxonomic annotation was available for originated from 163 unique species (**Figure 4.6 A)**. The three most abundant species identified were *Prevotella* sp. NE3005, *Prevotella ruminicola Ga686* and *Fibrobacter succinogenes*.

Respectively 49, 45 and 45 protein groups were identified from these three species (**Figure 4.6 B & Figure 4.6 C**).


A)
Number of protein groups identified from species
0
10
20
30
40
50
Prevotella sp. NE3005
Prevotella ruminicola Ga6B6
Fibrobacter succinogens
Prevotella ruminicola KHT3
Bacteroidales bacterium
Prevotella sp. P6B4
Gammaproteobacteria bacterium UBA1249…
Prevotellaceae bacterium MN60
Prevotella sp. TF2-5
Prevotella ruminicola strain BPI-162
Prevotella sp. TC2-28
Prevotella sp. BP1-148
Prevotella sp. AGR2160
Prevotella bryantii KHPX14
Prevotella brevis P6B11
Sarcina sp. DSM 11001
Clostridiales bacterium strain R-7
Succinivibrio dextrinosolvens strain ACV-10
Ruminobacter sp. RM87
Ruminococcaceae bacterium P7
Succinivibrio dextrinosolvens
Ruminococcus flavefaciens strain YRD2003
Piromyces sp. E2 v1.0
Olsenella umbonata strain DSM 22619
Neocallimastix californiae G1 v1.0
Clostridiales
Treponema bryantii strain B25
Succiniclasticum ruminis DSM11005
Succinatimonas sp. UBA1264 [genus]
Sharpea azabuensis DSM 18934
Ruminococcus albus strain AR67
Prevotella sp. BP1-145
Orpinomyces sp. strain C1A
Prevotellaceae bacterium KH2P17
Prevotella sp. KH2C16
Lachnospiraceae bacterium YSD2013
Bifidobacterium ruminantium DSM 6489
Bacteroides sp. Ga6A1
Treponema bryantii NK4A124
Sharpea azabuensis strain KH2P10
Ruminococcus flavefaciens ND2009
Ruminococcus flavefaciens MA2007
Ruminococcus bromii
Ruminococcaceae bacterium FB2012
Lachnospiraceae bacterium RM5
Lachnospiraceae bacterium NK4A144
Lachnobacterium bovis DSM 14045
Dorea sp. AGR2135
Clostridiales bacterium NK3B98
Butyrivibrio hungatei NK4A153
Bacteroides sp. Ga6A2
Bacteroides sp. AR20
Anaeromyces robustus v1.0

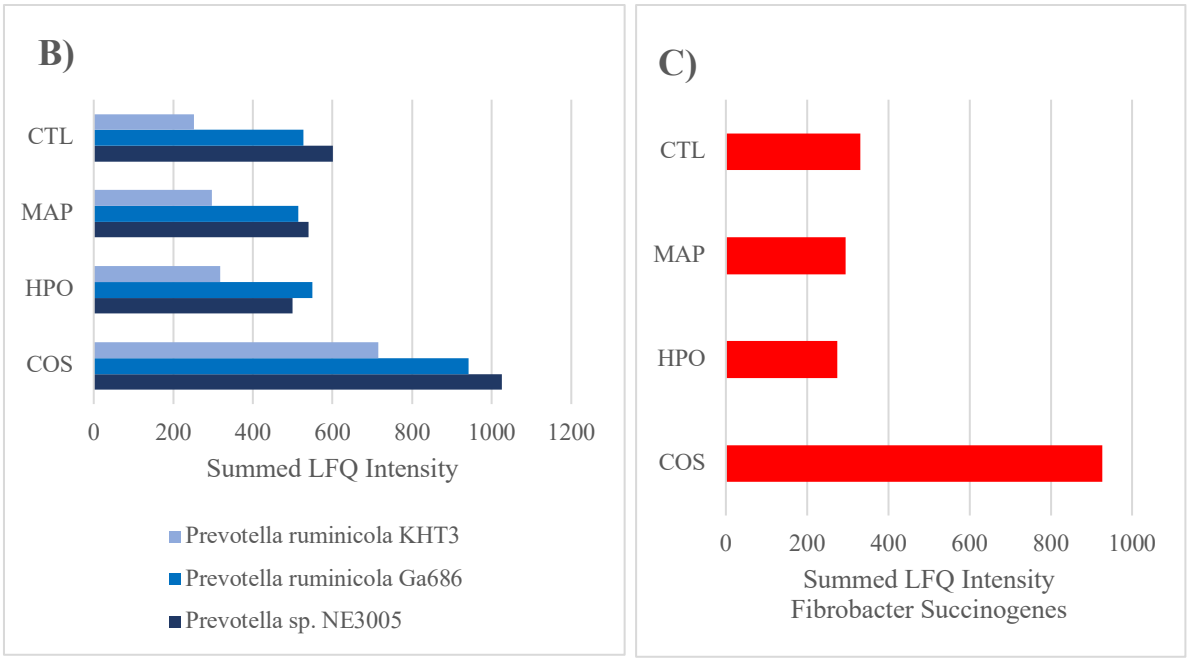


Figure 4.6. Distribution of the 52 most abundant species of the in total 163 species identified in MQrun 5. A) Abundance of 52 most abundant species identified from the 838 protein groups identified in MQrun5. Species with less than 3 identified protein groups were not included in this figure. Fungal species are marked in green, *Prevotella* species are marked in light blue and *Fibrobacter succinogenes* is marked in red. B) Summed LFQ intensities for all proteins across all biological replicates assigned to the three most abundant *Prevotella* strains; *Prevotella* sp. NE3005, *Prevotella ruminicola* Ga686 and *Prevotella ruminicola* KHT3. C) Summed LFQ intensities for all proteins across all biological replicates assigned to *Fibrobacter succinogenes*. Diet acronyms are as follows; CTL = basal control, no additional oil, HPO = basal concentrate supplemented with hydrogenated palm oil, MAP = basal concentrate supplemented with marine algae powder, COS = basal concentrate containing wheat starch and supplemented with corn oil.

4.4 Analysis of change in microbial composition

To examine the change in composition and metabolic function of the rumen microbial community, ANOVA analysis was run on the 838 protein groups identified in MQrun5, with p-value = 0,05 (**Figure 4.7** and **Figure 4.8**). The aim was to detect protein groups that were significantly different between the different feed among all 16 samples from cow animals. Samples originating from goats were not included, due to the low number of identified proteins (**Figure 4.1** and **Figure 4.2).** The results showed that most of the significant protein groups are expressed in samples where the cow had been fed the low-methane COS diet (**Figure 4.7).** In total, the ANOVA yielded 155 significant protein groups and the distribution of these groups are visualized as a heat map in **Figure 4.8A.** All 155 significant protein groups listed with InterPro annotation and taxonomic annotation in **Appendix A-2**. Notably, the hierarchical clustering of columns (i.e. cows; dendrogram in the top panel in **Figure 4.8A**) demonstrated

that samples from cows fed the low-methane COS diet clustered together, while samples from the other three diets seemed randomly distributed. In correlation with **Figure 4.7**, **Figure 4.8A** showed that samples where the cow had been fed COS diet also have the most identified protein groups, while we observe many missing hits for protein groups in the three other diets.

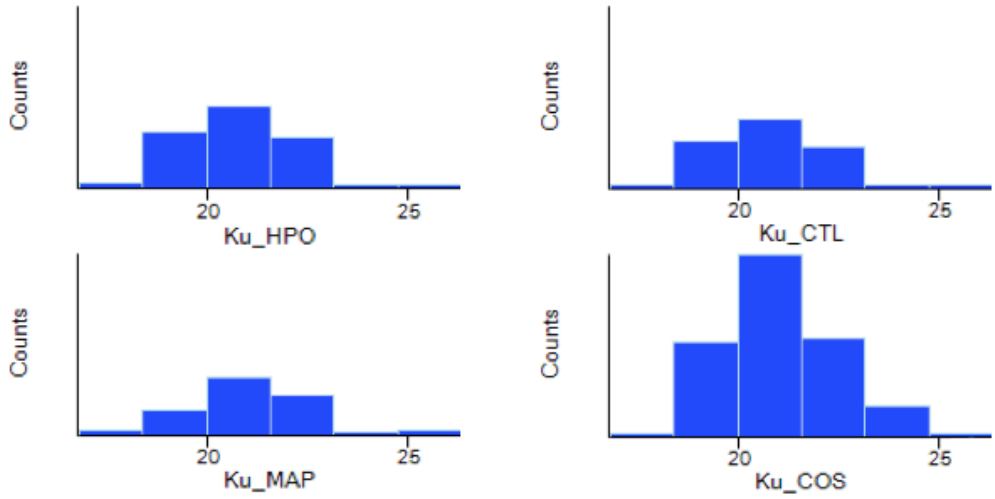


Figure 4.7 Histograms of counted significant protein groups for each diet for the ANOVA analysis, p-value=0.05. For this histogram the biological replicates were averaged. Samples from cow fed the low-methane COS diet have the most identified proteins, while the three other diets have many missing hits for protein groups. X-axis represent distribution of protein groups with low to higher LFQ intensities. Y-axis represent the number of identified protein groups with given LFQ intensity. Diet acronyms are as follows; CTL = basal control, no additional oil, HPO = basal concentrate supplemented with hydrogenated palm oil, MAP = basal concentrate supplemented with marine algae powder, COS = basal concentrate containing wheat starch and supplemented with corn oil.

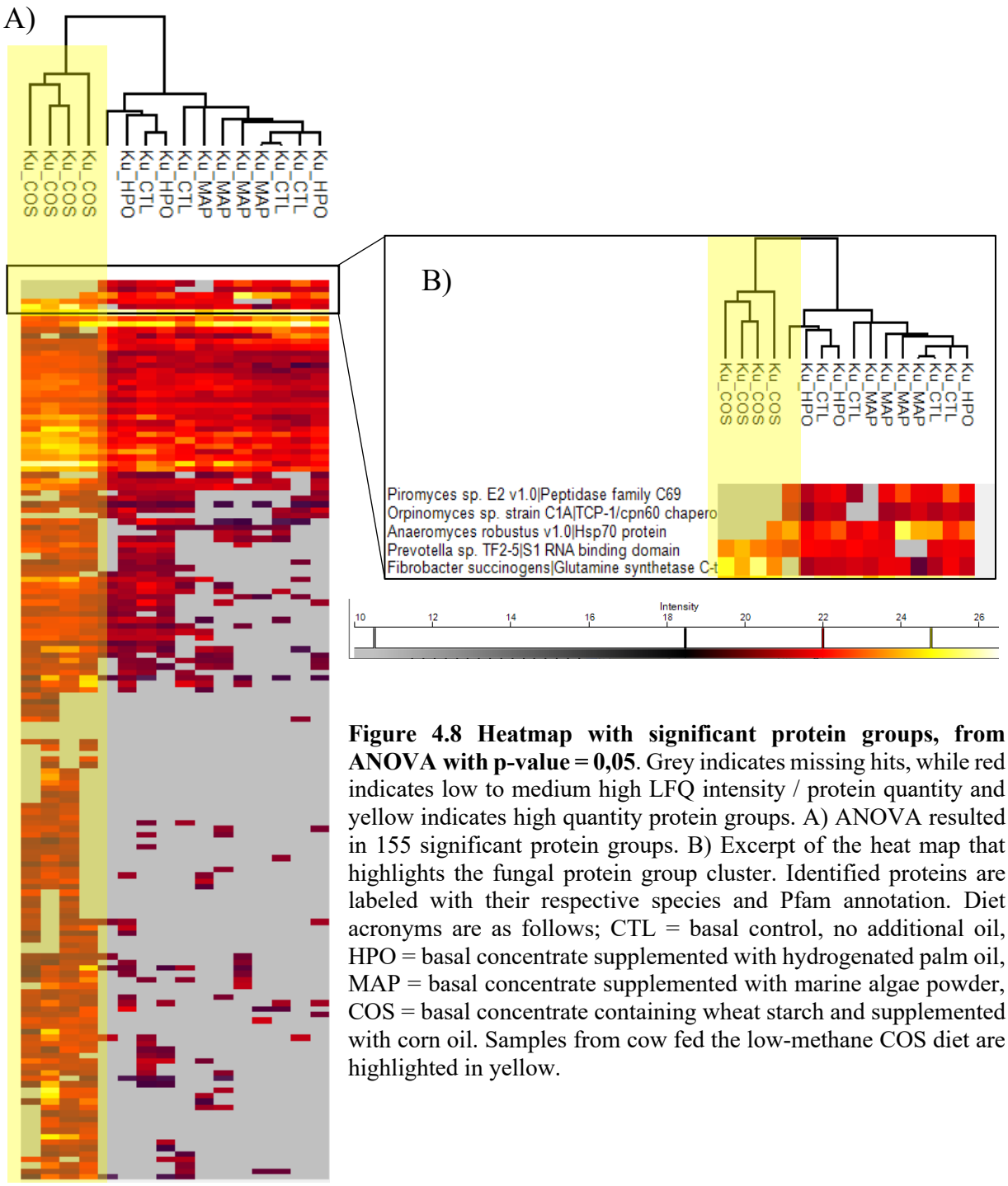


Figure 4.8 Heatmap with significant protein groups, from ANOVA with p-value = 0,05. Grey indicates missing hits, while red indicates low to medium high LFQ intensity / protein quantity and yellow indicates high quantity protein groups. A) ANOVA resulted in 155 significant protein groups. B) Excerpt of the heat map that highlights the fungal protein group cluster. Identified proteins are labeled with their respective species and Pfam annotation. Diet acronyms are as follows; CTL = basal control, no additional oil, HPO = basal concentrate supplemented with hydrogenated palm oil, MAP = basal concentrate supplemented with marine algae powder, COS = basal concentrate containing wheat starch and supplemented with corn oil. Samples from cow fed the low-methane COS diet are highlighted in yellow.

ANOVA analysis revealed a visibly increased amount of significant protein groups for samples from cow fed the low-methane COS diet compared to the three other diets. However, there is one area of the heatmap shown in **Figure 4.8A** that demonstrated missing hits (unidentified protein groups) for the samples from COS-fed cow (highlighted in yellow), compared to the three other diets, which all had identified proteins. This area of missing hits in COS are cropped

(highlighted in yellow) and shown in **Figure 4.8B** with identified proteins from species in this area, showed with their respective Pfam annotation. Interestingly, these proteins originate from respectively *Orpinomyces*, *Piromyces* and *Anaeromyces*, which are all cellulolytic and hydrogen producing fungal genera. In **Figure 4.8 A**, missing hits for MAP, HPO and CTL diets are observed where proteins have been identified from samples from cow fed the low-methane COS diet. Interestingly, many of these protein groups are affiliated to *F. succinogenes* (significant 8 protein groups) and *Prevotella* strains (significant 77 protein groups), respectively succinate and propionate producing species (**Appendix A-2**).

5 Discussion

The aim of this study was to quantify and identify proteins expressed by the rumen microbial community, in order to provide insight into how this microbial community changes when exposed to different diets, and how these changes can affect metabolic functions of the rumen microbiome. In this thesis, protein identification and change in microbial composition was analyzed through bottom-up proteomics sample preparation, high accuracy mass spectrometry analysis, database optimization and functional annotation of protein.

5.1 Sample preparation

5.1.1 Peptide clean-up with one-pot reaction strategy

Sample preparation in this thesis were conducted according to typical bottom-up proteomics workflow. After digestion of proteins into peptides by trypsin, the peptides were cleaned-up using ZipTips® Pipette Tips. As proteins and peptides can adhere to plastic, and to some extent also to glass ware, the number of tubes and vials (Christensen et al., 1978), during sample handling should be minimized. Here, we used a centrifugal approach, enabling efficient cleaning/elution of multiple ZipTips simultaneously while keeping the number of plastic surfaces at a minimum, where the latter contributes to an improved peptide recovery.

5.1.2 Separation of peptides by HPLC-gradient

The metaproteome of the rumen microbial community consist of a vast and complex mixture of proteins, which can be difficult to accurately identify and quantify. In order to ensure accurate and high-resolution protein quantification, the typical bottom-up proteomics workflow often include one or more fractionating step to separate proteins and/ or peptides (Issaq et al., 2002; Tanca et al., 2014), e.g. through SDS-PAGE or by HPLC-gradient. In this study we only used SDS-PAGE as a clean-up step for ("dirty", ruminal) samples by running the proteins only 1 cm into the gel. However, peptides were fractionated based on their hydrophobicity by reversed-phase chromatography. We used a gradient of aqueous and organic solvents in the HPLC prior to MS analysis. It is important to keep in mind that although fractioning samples are beneficial for accurate protein quantification, additional fractioning steps cause increased measuring time on MS and challenges regarding analytical reproducibility, as fractions are

prone to slight changes when analyzing multiple samples and comparing fractions (Cox et al., 2014; Tanca et al., 2014). However, this obstacle can be resolved in MaxQuant, by delayed normalization of fractions in order to compare them. By summing up LFQ intensities across fractions and using sophisticated normalization strategies, protein quantities can be determined accurately by employing global optimizing algorithms (Cox et al., 2014). In order to do so, the algorithms for LFQ scoring have the precondition that the majority of the proteome are not changing between samples (Cox et al., 2014), an assumption that also holds for the rumen microbiome, and hence the metaproteome samples analyzed in this thesis.

5.2 Protein sequence database selection and optimization

Compared to single cell-proteomics approaches, metaproteomics research has presented scientists with several challenges, notably due to the complexity and heterogeneity of samples (Muth et al., 2013). Microbial communities can consist of hundreds to thousands of different species, and thus encode up to several millions of proteins. Additionally, many of the species present in microbial samples may contain many closely related proteins, due to e.g. slight strain variations and horizontal gene transfer (Heyer et al., 2017; Muth et al., 2013). Accurate quantification and identification of the enormous amount of proteins in environmental metaproteomics samples is therefore highly dependent on a comprehensive and high-quality protein sequence database.

Proteomics protein identification is typically achieved via three different strategies; 1) de novo sequencing; referring to the interpretation of the amino acid sequence directly from MS spectra and subsequently identify the protein using a sequence-similarity search such as Basic Local Alignment Search Tool (BLAST). 2) Identification of peptide sequences using spectral libraries. 3) the most common approach in the "metaproteomics community", which is the identification of proteins via matching the experimental MS/MS spectra to theoretical fragmentation patterns of peptides found in an *in silico* peptide digestion from a user specific database (Muth et al., 2013). We used the third approach used to identify proteins for analyzing the rumen microbial community in this thesis. The protein database selection is essential and directly affects both (the number of) identified proteins, and their taxonomic assignment. As a result of the complexity of metaproteomic data (many MS/MS spectra and large databases), analyzing complex microbial communities require great computational effort in regard to efficient algorithms, in addition to increased memory power and processors (Heyer et al., 2017).

Moreover, protein identification will be inadequate if the selected protein database is missing entries for proteins present in the sampled microbiome due to novelty or limited insight into the microbial community composition (Heyer et al., 2017), or worse missing entire species. Slight changes in the DNA, such as single-nucleotide mutations, can alter the gene products, RNA and protein, and render identification process difficult and inaccurate when using a peptide sequence database. As mentioned previously, it is important that the protein database is constructed based on upstream metagenomics data acquired for sequencing the sample itself, or from a pseudo-metagenomics approach such as trimming large public repositories (e.g. the UniProt bacterial section) by 16S rRNA data (Muth et al., 2016; Tanca et al., 2016), or as done in this thesis, using a selection of genomes identified in other ruminal studies, where genomes and/ MAGs from metagenomes of similar microbial communities have been recruited to increase the specificity of the protein sequence databases. A sample-specific protein sequence database can help improve accurate identification, acquired from metagenomics sequencing and analyzing (i.e. assembly, binning, gene prediction) of the exact same sample, as this best reflect the state of the microbial community and the expression of proteins at the given time.

While a sample-specific metagenomic database is undoubtedly required for optimal protein identification rates, logistical and time-related reasons prevented the availability of a sample specific database for this study. Therefore, we endeavored to create alternative metagenomics databases using publicly available data and customize them as best as possible to fit our own metaproteomic data. Three rumen-specific databases were therefore constructed for protein identification, described in **Section 3.6.1**. A clear advantage of integrating metagenomic data with metaproteomics analysis is that it 1) provides protein sequences for all (most) proteins expected to be expressed by the microbial community and nothing more, i.e. sample-specific and a relatively small and concise database that should not give FDR issues. 2) It is usually complete. If using trimmed public repositories, the database may be incomplete due to lack of strain-specific variances or entire species may be missing as they are not present in the repository. 3) it can provide information about novel, uncultivated microorganism and their expressed proteins, thereby contributing to increased understanding of their metabolic functions in complex microbial communities. The protein sequence databases constructed in this thesis were based on rumen-associated MAGs reconstructed from two previously published metagenome studies conducted by respectively Hess et al. (2011) and Parks et al. (2017), in addition to the genomes of cultivated species from the Hungate 1000 database, and additional

methanogenic archaea genomes, as described in **Table 3.3**. Yet, it was not sample-specific per se and a level of incompleteness, i.e. missing species/ proteins were expected.

To ensure correct and accurate protein identification optimal databases should only contain proteins expressed by the species that are present in the community and should ideally not contain anything else. This was challenging as the exact phylogenetic composition is not known for the samples used for metaproteomics analysis in this study. Large databases, like RUDB1 that contains 425 559 protein entries, resulting a too little sample-similar database as it contains too large of a fraction of species not present in the sample. As shown in **Table 4.1**, only 5.61% of the MS/MS spectra are assigned to a peptide sequence in Run 1 using RUDB1, and subsequently lead to unsatisfactory protein identification. This was most likely due to the databases poor estimation of the microbial composition and proteins expressed in the rumen samples used in this study. In contrast, MQrun5, which used the RUDB3, had the highest MS/MS identification rate of 7.24% in MQruns1-5. When that said, for all MQruns combined MQrun6 had the highest MS/MS identification rate of 10.8% but the host protein check was conducted too late for its results to be further incorporated in this thesis. Host protein check is further discussed in **Section 5.2.1**. Compared to RUDB1, RUDB3 is a notably smaller database, with a reduction as of more than 100 000 protein entries compared to RUDB1. This indicates that RUDB3, as a more compact database, is the one out of the three that best reflects the proteins expressed by the rumen microbiome of the animals in this study. Downstream protein analysis was subsequently conducted with protein identified using this database.

MQrun2-4 were conducted using RUDB2 with varying FDRs and **Table 4.2** evidently showed the impact increased FDR can have on identified protein groups. Even though the number of identified MS/MS spectra and protein groups increases from MQrun2 to MQrun4, the number of identified false protein groups increased. The algorithms used for protein identification in proteomics search engines, such as MASCOT (Perkins et al., 1999), X!Tandem (Craig & Beavis, 2004) and Andromeda (Cox et al., 2011) used by MaxQuant in this study, suffers from the identification of false positives. A fraction of the identified PSMs will be false positives as the search engine attempts to match spectra that does not match the database (Muth et al., 2013). Strict control of FDR is therefore important in order to separate between true and false positive hits but using an FDR of 1% leads to only a low number of proteins identified. Some more proteins can be identified by increasing the FDR, but this also leads to more false protein identifications. It may indeed seem that large databases with a high level of similarity between

sequences, such as is common for metaproteomics databases, suffer from the target-decoy approach used for FDR control and identifies less than expected. Muth and colleagues have therefore, in order to increase the identification rate, suggested to use more than one search engine (Muth et al., 2013) or to use a two-step database search i.e. searching first a large database without reversed sequences or FDR control, and from this then generate a refined smaller database for a second target-decoy search (Jagtap et al., 2013). A variant of the latter was applied in the Galaxy sectioning approach (MQrun5) and may explain why this outperformed the standard searches in MQrun1-4. Also, since we are using MaxQuant for identification and quantification, it is not possible to integrate data from multiple search engines with the MaxQuant results without extensive bioinformatics expertise.

5.2.1 Quality control of protein quantification and identification with RUDB3

The sixth MaxQuant run described in **Section 3.6.1** and **Table 3.3** was conducted to check for the presence of host (cow; *Bos taurus*) proteins in the samples collected from cows. This MaxQuant run yielded a MS/MS identification rate of 10,8% and in total 2125 protein groups after filtration, where 198 protein groups originated from the host. This indicated three important aspects: 1) Very few host proteins are present in the rumen of the cows and 2) since the identification rate still is low, we have many MS/MS spectra that cannot be matched to a peptide sequence indicating that the RUDB3 database is most likely incomplete in terms of microbial composition. And 3) by using only half of the proteomics raw data (only cow) and including the host in the database, we can with the same parameters in MaxQuant, identify nearly the same amount of microbial proteins (742 here and 838 in MQrun5).

As previously mentioned, trypsin was used for the digestion of proteins into peptides, as it cleaves proteins after lysine and arginine, thus making each peptide easy to recognize. However, trypsin is likely to be influenced by other residues close the cleavage site and causing missed cleavages (Siepen et al., 2007). For all the MaxQuant runs, a maximum of one missed tryptic cleavage was allowed. Peptides with more than one missed cleavage were not evaluated and matched. For all MaxQuant runs, the samples were grouped and averaged for biological replicates. For a protein to be valid, the protein had to be identified in at least 2 of the four biological replicates. Ideally, a protein would only be considered valid if the protein was identified in all four replicates, increasing the confidence in the presence of the identified

protein. Because protein quantification was not achieved for all the replicates, such stringent criteria for protein validation would in turn lead to less identified protein groups as many proteins would be considered invalid. In order to identify as many protein groups as possible while maintaining an acceptable validation level, i.e. discard proteins that were identified by chance, it was required that a protein had to be identified in (at least) two of four replicates, i.e. half of the replicates.

5.3 Protein identification and microbial composition reflected in database

Despite the sub-optimality given by the lack of sample-specific database and the FDR issues discussed above, 838 microbial proteins were identified, and the replicates indicated comparable data (**Figure 4.3)**. The relative quantification showed to be reproducible between biological replicates, showing Pearson correlations (R) between LFQ intensities for samples from cow fed the low-methane COS diet varying from 0.417 and 0.812, which were slightly lower, but similar to those observed in other studies (Arntzen et al., 2015; Hagen et al., 2017). The highest Pearson correlation between Cow 1 and Cow 3 fed the COS diet at 0.812 (**Figure 4.3**), yielding a coefficient of determination (R^2) of 0.66, i.e. 66% of the variance in protein groups are shared by these two replicates. Pearson correlations from samples from cow are shown in **Appendix A-1**. Additionally, **Figure 4.2** showed that expressed proteins by one animal group were not identified in the other animal group (i.e. cow vs. goat), indicating that the microbial composition in the two animal groups may differ from each other even though similar feed were administered. Moreover, the formation of the two protein profile clusters from samples obtained from cow either or goat, reaffirmed that protein quantification patterns are similar within the animal groups, but differ between them (**Figure 4.2**). Samples from goat had visibly fewer identified protein groups, suggesting that the database was not optimal for samples from goat rumen. Although the RUDB3 was the database that identified the highest amount protein groups, the database identified fewer protein groups for generally all the samples from goat compared to the amount of protein groups identified for samples from cow. Even though Seshadri and colleagues revealed a genomic coverage of 75% of the known genera from the rumen when they published the Hungate 1000 genome project, the complete core microbial composition of the rumen microbiome is still unknown (Denman et al., 2018; Seshadri et al., 2018). Moreover, both the MAGs from Hess et al., and Parks et al., as well as

those from the Hungate 1000 database are heavily dominated by representatives from the cow rumen, meaning that the goat rumen is still very much genomically under-sampled (Denman et al., 2018; Kim et al., 2011). In this respect, the construction of a non-sample-specific protein sequence databases in this study, likely have biases resulting from the public databases that are skewed for cow, and as such, they cannot reflect the true microbial diversity found in the goat rumen.

Given the very low identification of proteins from the goat rumen microbiome, the focus was refined on detailed taxonomic and function annotation of proteins identified from the cow samples (**Figure 4.7** and **Figure 4.8**). Because RUDB3 was not sample-specific, i.e. consisted of only expressed proteins by present microorganism, but based on known genome sequence information, we can observe a similar identification pattern here as was observed for goats i.e. the lack of protein identifications for some of the diets. Interestingly, it was observed that the ANOVA analysis identified proteins resulted in samples from cows fed the COS diet clustering together, while the clustering for the three other samples seemed more random. This may suggest that the microbes included in RUDB3 reflects the microbial community in COS-fed animals more precisely than for the three other diets. In addition to co-clustering, **Figure 4.7** illustrated that using the RUDB3 database identified more protein groups in samples from cow fed the COS diet compared to the other three diets. This suggest that the microbial composition of samples from cow fed the COS diet shifted in response to the feed to a state that was different from the other three diets. This effect in varying protein identification between diets may indicate a change in microbial composition in the rumen microbial community in response to diet, and that this change in composition and protein expression can cause different metabolic functions to be carried out by the rumen microbiome. However, and due to the composition of RUDB3, our observations indicate the loss off certain species/proteins/functions for the three diets compared to COS diet, but we are not able to identify the gains, i.e. in the heatmap there is no region (except one; **Figure 4.8 B** and will be discussed further down) where proteins/species/functions are identified in CTL, HPO and MAP diets, that are not present in the low-methane COS diet. It could for example be that certain microbiome functions are maintained under the other feed, but by other microbes not present in our database; hence we cannot capture this change.

5.4 Metabolic effect of microbial change as a result of diet

The functional annotation of the database was done with InterProScan5 version 5.32-71.0 and was available for 396 of the 716 protein groups where taxonomic annotation was available. InterProScan combines the different search applications in InterPro for prediction of protein function (Mitchell et al., 2018). InterPro is a large database, compromised of multiple functional databases such as Pfam, Prosite and TIGRFAMs, which collectively offer predictive information about protein function and protein families, in addition to domain and site information. By combining these different search applications InterProScan predicts protein function and summarize their outputs. InterProScan also identify a corresponding InterPro entry and other database annotations, e.g. Gene Ontology (GO) for the identified protein (Jones et al., 2014). As InterPro consists of many specific databases, InterProScan can obtain more annotations that running a single database search in one of these databases, i.e. Pfam (El-Gebali et al., 2018). In this study, all of the database annotations were added to annotation containers in the Perseus matrix, enabling functional properties and metabolic pathway annotation. Even though Pfam has a large coverage of proteins and protein families, we experienced that InterPro were able to provide functional annotations to more protein groups than Pfam in the Perseus matrix. Hence functional annotation of 155 significant protein groups from the ANOVA, showed in **Appendix A-2**, were annotated with their InterPro annotations. As shown in **Figure 4.4,** 99 of the 396 protein groups were assigned to the GO term oxidation-reduction processes, while carbohydrate metabolic processes, in addition to glycolytic processes and gluconeogenesis are abundantly represented as dominating protein function GO term categories with respectively 25, 31 and 15 assigned protein groups, hence reflecting the carbohydrate degrading and fermenting functional properties of the rumen microbiome. This observed and identified functional profile can also argue for that while RUDB3 were not able to reflect the complete metaproteomic contribution of all the samples submitted for MS/MS analysis, RUDB3 were able to reflect key metabolic functions in the rumen microbiome.

All of the 155 significant proteins from the ANOVA analysis were of particular interest, as these protein groups were expressed significantly different in one diet compared to the other three. These results explicitly demonstrated how the metabolic functions of the microbial populations in the rumen were affected when exposed to different diets with varying content of

lipid and complex carbohydrates. Most interestingly, **Figure 4.7** showed that most of the significant protein groups from the ANOVA were expressed in the samples where the cow had been fed the COS diet, which also demonstrated a substantial decrease in measured enteric methane levels compared to the cows fed CTL, MAP and HPO diets (**Table 4.1**). The ANOVA results highlighted a clear change in some proteins (**Figure 4.8 A**) between diets, especially for fungal proteins (highlighted in **Figure 4.8 B**) that were present in all diets, except the low-methane COS diet. This not only reaffirms our earlier statement that RUDB3 best reflects the microbiome in the rumen of COS-fed cows, but that the COS diet is "manipulating" the rumen microbiota and creating a change in the rumen microbiome that is (likely) correlated with lower methane production observed in the original study by INRA by Fougère et al. (2018).

Although RUDB3 included 20 non-redundant methanogens, only three protein groups from methanogenic archaea were identified amongst the 716 protein groups that taxonomic annotation was available for (**Figure 4.5 A)**. The identified protein groups were annotated to originate from *Methanobrevibacter* and *Methanosarcina,* common methanogens whose functional profile in the rumen microbiome is well-known and acknowledged (Marvin-Sikkema et al., 1993; Moran, 2005; Roque et al., 2019). As stated above, methane production for the animals used to provide samples in this study have been measured in the metadata provided by INRA. It would therefore be expected that more protein groups assigned to methanogens were identified in this study. While RUDB3 consisted of as many of the known rumen methanogens as possible, these results indicate the possibility that only a minor fraction of the actual methane producing population in the ruminant animals used in this study are known (Patra et al., 2017). Such a scenario could be addressed in future studies that develop and utilize a sample specific database to improve protein identification rates and greater evaluate the change in the methane producing population in the rumen directly.

While this analysis only detected few proteins affiliated with methanogens, an interesting pattern was observed that could provide explanations as to the difference in methane levels observed in COS-fed animals and the other diet groups. The results illustrated a decreased detection of fungal proteins in the samples from cow fed the low-methane COS diet (**Figure 4.8 A)**, indicating a decrease in fungal metabolic activity, i.e. cellulolytic activity and hydrogen production. The absence of fungal hydrogen producing species, such as *Piromyces*, *Anaeromyces* and *Orpinomyces*, may affect the methanogenic population of the rumen. Therefore, a decrease in available hydrogen in the rumen could theoretically lead to a decrease

in methane production, which is in coherence with Fougère et al. (2018) and their measurement of decreased methane production in animals fed the high starch COS diet. In contrast to lower fungal activity, **Figure 4.6 B** and **Figure 4.6 C** illustrated increased detection of protein groups expressed in COS-fed cows that were affiliated to different *Prevotella strains* and *Fibrobacter succinogenes*, which were found to be the most abundant species. As mentioned, *F. succinogenes* is a cellulose degrading bacteria that uses hydrogen to produce succinate, that subsequently can be converted to propionate via the other succinate-consuming microbiota in the rumen (Scheifinger & Wolin, 1973). Similarly, the fiber-degradation *Prevotella* are linked to propionate production in the rumen (Chen et al., 2017), a pathway that also requires hydrogen. The increase in identified proteins affiliated with species that redirect hydrogen to volatile fatty acid (VFA) production (succinate and/or propionate) in the rumen, suggests that less hydrogen is released to the total "hydrogen pool" in the rumen, that would otherwise be converted into methane by methanogens. The increase in VFA productions is beneficial both for the ruminant and the environment, as VFAs can be further utilized in energy metabolism and less methane is produced and emitted.

5.5 Future research and concluding remarks

The aim of this study was to quantify and identify proteins in the rumen microbial community, in order to better understand how different diets can affect and change the microbial population in the rumen and the metabolic functions that are carried out. The metaproteome of the rumen was analyzed from both cows and goats that were fed contrasting diets differing in lipid content, which ultimately caused variations in the enteric methane being released. As a part of the metaproteomic analysis, a sample specific protein sequence database was not available for the samples provided for this study, hence three different protein sequence databases were constructed to best estimate the microbial genomic content and metaproteomic potential of the rumen. The databases consisted of (meta)genomic data from both MAGs and cultivated genomes identified in the rumen microbiome. By combining (meta)genomic information in the metaproteomics analysis (i.e. multi-omics), we were able to assess the functional expression of the rumen microbiome in response to contrasting dietary conditions.

Rumen fluid, from goats and cows fed four different diets supplemented with different lipid sources and complex carbohydrates, was prepared for high accuracy mass spectrometry

analysis in typical bottom-up proteomics approach and quantified and analyzed using the proteomics software MaxQuant and Perseus. Of the three databases that was tested, RUDB3 was the database that identified the most MS/MS spectra and hence reflected the rumen microbial population best. Because RUDB3 was not entirely sample-specific, RUDB3 was less optimal for the samples from goat rumen than the samples from cow rumen as there exists fra less metagenomic data that is representative of the goat rumen microbiome. However, using RUDB3, we were able to detect prominent changes in microbial composition when the cow rumen was subjected to different diets. Metadata from a previous study conducted by the sample provider (INRA) had already showed that animals fed a diet supplemented with corn oil and wheat starch (COS), decreased the production of enteric methane in the rumen. Using the RUDB3 database, we identified the highest number of proteins from COW-fed cow rumen samples, which upon closer examination indicated a redirection of hydrogen from methane production to volatile fatty acid production by primarily succinate-producing *Fibrobacter succinogenes* and propionate-producing *Prevotella* strains. In addition, the COS diet seemingly affected the abundance of hydrogen-producing fungal species, which we speculate may be a chief reason that low levels of hydrogen-utilizing methanogenic archaea were detected and less methane production in the rumen was observed in COS-fed cows.

For future research, comparing the results presented in this study conducted with a custom database derived from publicly available data against results from a sample-specific database would be interesting and of high value. A sample-specific database would provide more protein identifications and should be able to provide in-depth information in regard to metabolic function carried out by the rumen microbiome not only when subjected to the high starch, low methane COS diet, but also the three other diets from this study. In addition, these results have shown that the host organisms (cow and goat) should be included in the database used. Lastly, the integration of culture independent techniques, such as metagenomics, with functional -omics tool such as metaproteomics, in further studies of the complex rumen microbiome, enables a greater understanding of the rumen metabolic functions, its microbial populations and their interactions, and how the rumen microbiome can be manipulated by diets to decrease greenhouse gas emissions while ensuring animal health.

6 References

Aebersold, R. & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, 537 (7620): 347.

Agarwal, U., Hu, Q., Baldwin, R. & Bequette, B. (2015). Role of rumen butyrate in regulation of nitrogen utilization and urea nitrogen kinetics in growing sheep. *Journal of animal science*, 93 (5): 2382-2390.

Annison, F., Bramley, E., Browning, G., Cusack, P., Farquharson, B., Lean, I., Little, S. & Nandapi, D. (2007). Ruminal Acidosis–aetiopathogenesis, prevention and treatment. *A review for veterinarians and nutritional professionals*: 7-43.

Antanaitis, R., Zˇilaitis, V., Juozaitiene, V. & Stosˇkus, R. (2016). Usefulness of acidity and temperature of the rumen and abomasum in diagnosing SARA in dairy cows after calving. *Polish Journal of Veterinary Sciences*, Vol. 19 (No. 3 (2016)): 553–558. doi: DOI 10.1515/pjvs-2016-0069.

Arntzen, M. Ø., Karlskås, I. L., Skaugen, M., Eijsink, V. G. H. & Mathiesen, G. (2015). Proteomic Investigation of the Response of Enterococcus faecalis V583 when Cultivated in Urine. *PLOS ONE*, 10 (4): e0126694. doi: 10.1371/journal.pone.0126694.

Bashiardes, S., Zilberman-Schapira, G. & Elinav, E. (2016). Use of metatranscriptomics in microbiome research. *Bioinformatics and biology insights*, 10: BBI. S34610.

Beauchemin, K. & McGinn, S. (2006). Methane emissions from beef cattle: Effects of fumaric acid, essential oil, and canola oil. *Journal of Animal Science*, 84 (6): 1489-1496.

Bio-Rad. (2019). *DC Protein Assay Instruction Manual*. PDF. http://www.bio-rad.com/webroot/web/pdf/lsr/literature/LIT448.pdf (lest 23/3).

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*, 72 (1-2): 248-254.

Chen, T., Long, W., Zhang, C., Liu, S., Zhao, L. & Hamaker, B. R. (2017). Fiber-utilizing capacity varies in Prevotella-versus Bacteroides-dominated gut microbiota. *Scientific reports*, 7 (1): 2594.

Chilliard, Y. (1993). Dietary fat and adipose tissue metabolism in ruminants, pigs, and rodents: A review. *Journal of Dairy Science*, 76 (12): 3897-3931.

Christensen, P., Johansson, A. & Nielsen, V. (1978). Quantitation of protein adsorbance to glass and plastics: Investigation of a new tube with low adherence. *Journal of immunological methods*, 23 (1-2): 23-28.

Cox, J. & Mann, M. (2007). Is proteomics the new genomics? *Cell*, 130 (3): 395-398.

Cox, J. & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26: 1367. doi: 10.1038/nbt.1511

https://www.nature.com/articles/nbt.1511#supplementary-information.

Cox, J. & Mann, M. (2011). Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry*, 80: 273-299.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V. & Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research*, 10 (4): 1794-1805.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N. & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics*, 13 (9): 2513-2526.

Craig, R. & Beavis, R. C. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, 20 (9): 1466-1467.

Cravatt, B. F., Simon, G. M. & Yates Iii, J. R. (2007). The biological impact of mass-spectrometry-based proteomics. *Nature*, 450 (7172): 991.
De Beni Arrigoni, M., Martins, C. L. & Factori, M. A. (2016). Lipid Metabolism in the Rumen. I: Millen, D. D., De Beni Arrigoni, M. & Lauritano Pacheco, R. D. (red.) *Rumenology*, s. 103-126. Cham: Springer International Publishing.
Denman, S., Morgavi, D. & McSweeney, C. (2018). The application of omics to rumen microbiota function. *animal*, 12 (s2): s233-s245.
Diaz-Viraque, F., Pita, S., Greif, G., de Souza, R. d. C. M., Iraola, G. & Robello, C. (2018). Nanopore sequencing significantly improves genome assembly of the eukaryotic protozoan parasite Trypanosoma cruzi. *bioRxiv*: 489534.
Dijkstra, J. (1994). Production and absorption of volatile fatty acids in the rumen. *Livestock Production Science*, 39 (1): 61-69.
Ding, S.-Y., Rincon, M. T., Lamed, R., Martin, J. C., McCrae, S. I., Aurilia, V., Shoham, Y., Bayer, E. A. & Flint, H. J. (2001). Cellulosomal Scaffoldin-Like Proteins fromRuminococcus flavefaciens. *Journal of Bacteriology*, 183 (6): 1945-1953.
Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., et al. (2014). *IPCC, 2014: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*: Cambridge University Press.
El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A. & Smart, A. (2018). The Pfam protein families database in 2019. *Nucleic acids research*, 47 (D1): D427-D432.
EPA. (2019). *Inventory of Greenhouse Gas Emissions and Sink: 1990-2017*: EPA - United States Environmetal Protection Agency.
Fougère, H., Delavaud, C. & Bernard, L. (2018). Diets supplemented with starch and corn oil, marine algae, or hydrogenated palm oil differentially modulate milk fat secretion and composition in cows and goats: A comparative study. *Journal of dairy science*, 101 (9): 8429-8445.
Fougère, H. & Bernard, L. (2019). Effect of diets supplemented with starch and corn oil, marine algae, or hydrogenated palm oil on mammary lipogenic gene expression in cows and goats: A comparative study. *Journal of dairy science*, 102 (1): 768-779.
Fraser, M. (2017). *What is Interproscan?* https://github.com/ebi-pf-team/interproscan/wiki. Tilgjengelig fra: https://github.com/ebi-pf-team/interproscan/wiki (lest 26/4).
Gadeyne, F., De Neve, N., Vlaeminck, B. & Fievez, V. (2017). State of the art in rumen lipid protection technologies and emerging interfacial protein cross-linking methods. *European Journal of Lipid Science and Technology*, 119 (5): 1600345.
Gruninger, R. J., Puniya, A. K., Callaghan, T. M., Edwards, J. E., Youssef, N., Dagar, S. S., Fliegerova, K., Griffith, G. W., Forster, R. & Tsang, A. (2014). Anaerobic fungi (phylum Neocallimastigomycota): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS microbiology ecology*, 90 (1): 1-17.
Hagen, L. H., Frank, J. A., Zamanzadeh, M., Eijsink, V. G., Pope, P. B., Horn, S. J. & Arntzen, M. Ø. (2017). Quantitative metaproteomics highlight the metabolic contributions of uncultured phylotypes in a thermophilic anaerobic digester. *Appl. Environ. Microbiol.*, 83 (2): e01955-16.
Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology*, 5 (10): R245-R249.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Collaborators, G. R. C., Abecia, L., Angarita, E., Aravena, P. & Arenas, G. N. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific reports*, 5: 14567.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D. S., Chen, F. & Zhang, T. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 331 (6016): 463-467.

Heyer, R., Schallert, K., Zoun, R., Becher, B., Saake, G. & Benndorf, D. (2017). Challenges and perspectives of metaproteomic data analysis. *Journal of biotechnology*, 261: 24-36.

Horn, S. J., Vaaje-Kolstad, G., Westereng, B. & Eijsink, V. (2012). Novel enzymes for the degradation of cellulose. *Biotechnology for Biofuels*, 5 (1): 45. doi: 10.1186/1754-6834-5-45.

Hristov, A., Oh, J., Firkins, J., Dijkstra, J., Kebreab, E., Waghorn, G., Makkar, H., Adesogan, A., Yang, W. & Lee, C. (2013). Special topics—Mitigation of methane and nitrous oxide emissions from animal operations: I. A review of enteric methane mitigation options. *Journal of animal science*, 91 (11): 5045-5069.

Issaq, H. J., Conrads, T. P., Janini, G. M. & Veenstra, T. D. (2002). Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis*, 23 (17): 3048-3061.

Jagtap, P., Goslinga, J., Kooren, J. A., McGowan, T., Wroblewski, M. S., Seymour, S. L. & Griffin, T. J. (2013). A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics*, 13 (8): 1352-1357.

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

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A. & Nuka, G. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30 (9): 1236-1240.

Keiblinger, K. M., Fuchs, S., Zechmeister-Boltenstern, S. & Riedel, K. (2016). Soil and leaf litter metaproteomics—a brief guideline from sampling to understanding. *FEMS microbiology ecology*, 92 (11).

Kim, M., Morrison, M. & Yu, Z. (2011). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS microbiology ecology*, 76 (1): 49-63.

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

Kruger, N. J. (2009). The Bradford method for protein quantitation. I: *The protein protocols handbook*, s. 15-21: Springer.

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

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2013). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research*, 42 (D1): D490-D495.

Lourenço, M., Ramos-Morales, E. & Wallace, R. (2010). The role of microbes in rumen lipolysis and biohydrogenation and their manipulation. *Animal*, 4 (7): 1008-1023.

Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *Journal of biological chemistry*, 193: 265-275.

Machado, L., Magnusson, M., Paul, N. A., Kinley, R., de Nys, R. & Tomkins, N. (2016). Dose-response effects of Asparagopsis taxiformis and Oedogonium sp. on in vitro fermentation and methane production. *Journal of Applied Phycology*, 28 (2): 1443-1452. doi: 10.1007/s10811-015-0639-9.

Marvin-Sikkema, F., Richardson, A., Stewart, C., Gottschal, J. & Prins, R. (1990). Influence of hydrogen-consuming bacteria on cellulose degradation by anaerobic fungi. *Appl. Environ. Microbiol.*, 56 (12): 3793-3797.

Marvin-Sikkema, F. D., Rees, E., Kraak, M. N., Gottschal, J. C. & Prins, R. A. (1993). Influence of metronidazole, CO, CO2, and methanogens on the fermentative metabolism of the anaerobic fungus Neocallimastix sp. strain L2. *Appl. Environ. Microbiol.*, 59 (8): 2678-2683.

Medical & Biological Laboratories CO., L. (2017). *The principle and method of polyacrylamide gel electrophoresis (SDS-PAGE)*. Web page. Tilgjengelig fra: https://www.mblbio.com/bio/g/support/method/sds-page.html (lest 8/4).

Merck Millipore. (2019). *ZipTip® Pipette Tips: Concentrating and purifying samples for MALDI-ToF MS*: Merck. Tilgjengelig fra: http://www.merckmillipore.com/NO/en/product/ZipTip-Pipette-Tips,MM_NF-C5737?ReferrerURL=https%3A%2F%2Fwww.google.com%2F#overview (lest 8/4).

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H.-Y., El-Gebali, S. & Fraser, M. I. (2018). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic acids research*, 47 (D1): D351-D360.

Moraïs, S., Morag, E., Barak, Y., Goldman, D., Hadar, Y., Lamed, R., Shoham, Y., Wilson, D. B. & Bayer, E. A. (2012). Deconstruction of Lignocellulose into Soluble Sugars by Native and Designer Cellulosomes. *mBio*, 3 (6): e00508-12. doi: 10.1128/mBio.00508-12.

Moran, J. (2005). *Tropical dairy farming: feeding management for small holder dairy farmers in the humid tropics*: Csiro publishing.

MP Biomedicals. (2018). *FastPrep® Instruments, Bead Beating Instruments for Sample Preparation*. pdf. https://eu.mpbio.com/pub/media/productattachment/LS012019-EN-FastPrep-Instruments-Brochure.pdf (lest 23/03).

Mullis, K. B. & Faloona, F. A. (1987). [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. I: b. 155 *Methods in enzymology*, s. 335-350: Elsevier.

Muth, T., Benndorf, D., Reichl, U., Rapp, E. & Martens, L. (2013). Searching for a needle in a stack of needles: challenges in metaproteomics data analysis. *Molecular BioSystems*, 9 (4): 578-585.

Muth, T., Renard, B. Y. & Martens, L. (2016). Metaproteomic data analysis at a glance: advances in computational microbial community proteomics. *Expert review of proteomics*, 13 (8): 757-769.

NCBI. (2019). *Dithiothreitol, CID=446094*. Wed Page. PubChem Database: National Center for Biotechnology Information. Tilgjengelig fra: https://pubchem.ncbi.nlm.nih.gov/compound/446094 (lest 8/4).

Naas, A. E., Mackenzie, A. K., Mravec, J., Schückel, J., Willats, W. G. T., Eijsink, V. G. H. & Pope, P. B. (2014). Do Rumen Bacteroidetes Utilize an Alternative Mechanism for Cellulose Degradation? *mBio*, 5 (4): e01401-14. doi: 10.1128/mBio.01401-14.

Naas, A. E., Norges miljø- og biovitenskapelige universitet Fakultet for kjemi, b. o. m. & Norges miljø- og biovitenskapelige, u. (2017). *Investigating the cellulolytic potential of the cow rumen metagenome = Undersøkelser av det cellulolytiske potensialet i kuvommas metagenom*. Ås: Norwegian University of Life Sciences, Faculty of Chemistry, Biotechnology and Food Science.

Olivier, J. G., Schure, K. & Peters, J. (2017). Trends in global CO2 and total greenhouse gas emissions. *PBL Netherlands Environmental Assessment Agency*: 5.

Owens, F. N. & Basalan, M. (2016). Ruminal Fermentation. I: Millen, D. D., De Beni Arrigoni, M. & Lauritano Pacheco, R. D. (red.) *Rumenology*, s. 63-102. Cham: Springer International Publishing.

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P. & Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature microbiology*, 2 (11): 1533.

Patra, A., Park, T., Kim, M. & Yu, Z. (2017). Rumen methanogens and mitigation of methane emission by anti-methanogenic compounds and substances. *Journal of Animal Science and Biotechnology*, 8 (1): 13. doi: 10.1186/s40104-017-0145-9.

Patti, G., Yanes, O. & Siuzdak, G. (2012). Metabolomics: the apogee of the omics trilogy: Innovation. *Nat Rev Mol Cell Biol*, 13: 263-269.

Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS: An International Journal*, 20 (18): 3551-3567.

Petri, R. M., Schwaiger, T., Penner, G. B., Beauchemin, K. A., Forster, R. J., McKinnon, J. J. & McAllister, T. A. (2014). Characterization of the Core Rumen Microbiome in Cattle during Transition from Forage to Concentrate as Well as during and after an Acidotic Challenge. *PLOS ONE*, 8 (12): e83424. doi: 10.1371/journal.pone.0083424.

Redmile-Gordon, M., Armenise, E., White, R., Hirsch, P. & Goulding, K. (2013). A comparison of two colorimetric assays, based upon Lowry and Bradford techniques, to estimate total protein in soil extracts. *Soil Biology and Biochemistry*, 67: 166-173.

Resing, K. A. & Ahn, N. G. (2005). Proteomics strategies for protein identification. *FEBS letters*, 579 (4): 885-889.

Roque, B. M., Brooke, C. G., Ladau, J., Polley, T., Marsh, L. J., Najafi, N., Pandey, P., Singh, L., Salwen, J. K., Eloe-Fadrosh, E., et al. (2019). Effect of the macroalgae Asparagopsis taxiformis on methane production and rumen microbiome assemblage. *Animal Microbiome*, 1 (1): 3. doi: 10.1186/s42523-019-0004-4.

Sanger, F. & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*, 94 (3): 441-448.

Scheifinger, C. C. & Wolin, M. J. (1973). Propionate formation from cellulose and soluble sugars by combined cultures of Bacteroides succinogenes and Selenomonas ruminantium. *Appl. Environ. Microbiol.*, 26 (5): 789-795.

Scheller, H. V. & Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology*, 61 (1): 263-289. doi: 10.1146/annurev-arplant-042809-112315.

Schwarz, W. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. *Applied microbiology and biotechnology*, 56 (5-6): 634-649.

Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Hadjithomas, M. & Varghese, N. J. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nature biotechnology*, 36 (4): 359.

Siepen, J. A., Keevil, E.-J., Knight, D. & Hubbard, S. J. (2007). Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *Journal of proteome research*, 6 (1): 399-408.

Sigma. (2001). *Product Information: Iodoacetamide*: Sigma-Aldrich.

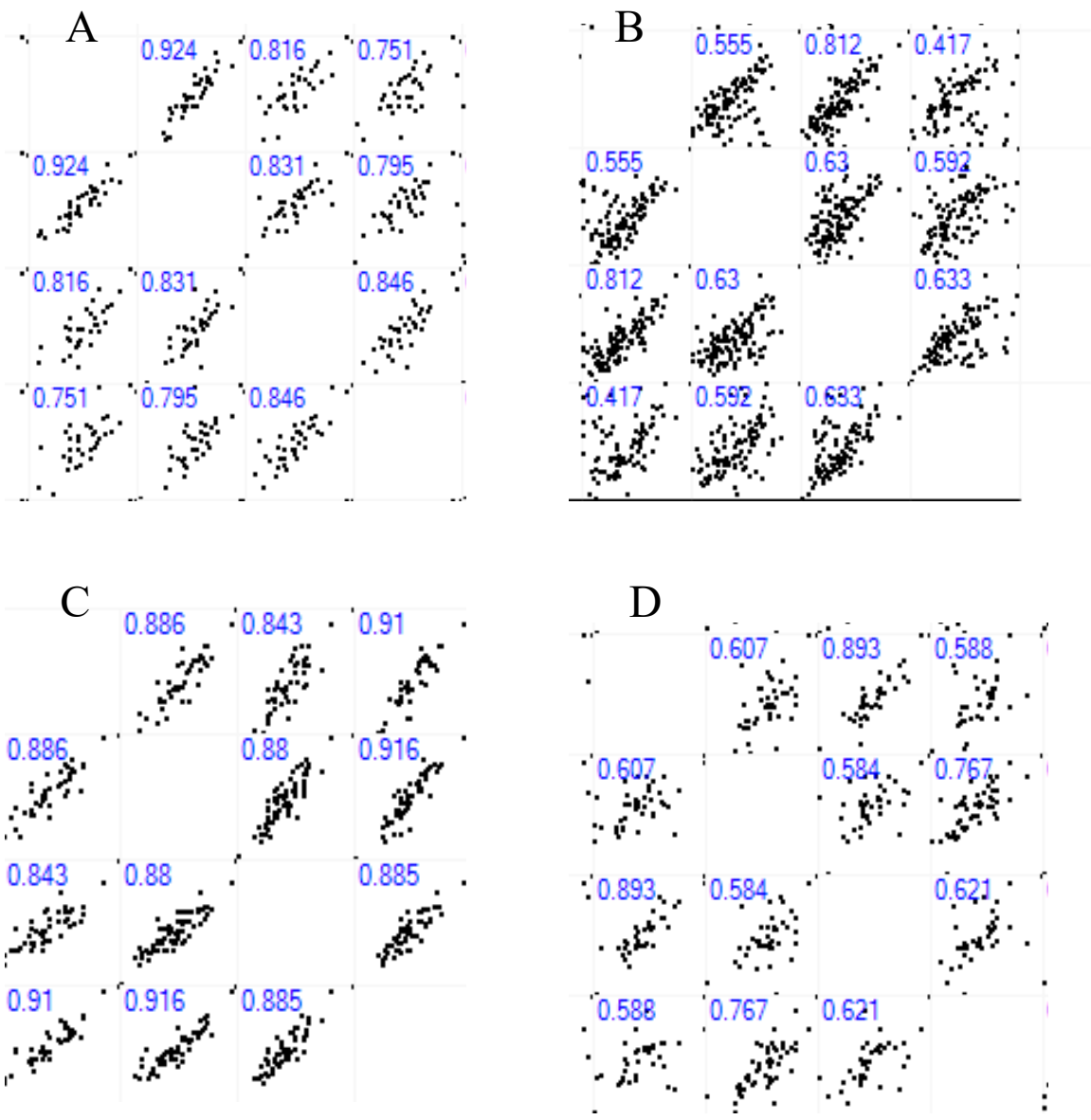
Stewart, R. D., Auffret, M. D., Warr, A., Wiser, A. H., Press, M. O., Langford, K. W., Liachko, I., Snelling, T. J., Dewhurst, R. J. & Walker, A. W. (2018). Assembly of 913 microbial

genomes from metagenomic sequencing of the cow rumen. *Nature communications*, 9 (1): 870.
Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. & (eds.), P. M. M. (2013). *IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.
Tanca, A., Palomba, A., Pisanu, S., Deligios, M., Fraumene, C., Manghina, V., Pagnozzi, D., Addis, M. F. & Uzzau, S. (2014). A straightforward and efficient analytical pipeline for metaproteome characterization. *Microbiome*, 2 (1): 49.
Tanca, A., Palomba, A., Fraumene, C., Pagnozzi, D., Manghina, V., Deligios, M., Muth, T., Rapp, E., Martens, L. & Addis, M. F. (2016). The impact of sequence database choice on metaproteomic results in gut microbiota studies. *Microbiome*, 4 (1): 51.
Tapio, I., Snelling, T. J., Strozzi, F. & Wallace, R. J. (2017). The ruminal microbiome associated with methane emissions from ruminant livestock. *Journal of animal science and biotechnology*, 8 (1): 7.
Thermo Fischer Scientific. (2019). *NuPAGE™ Sample Reducing Agent (10X)*. Web page. https://www.thermofisher.com: Thermo Fischer Scientific. Tilgjengelig fra: https://www.thermofisher.com/order/catalog/product/NP0004 (lest 8/4).
Thiede, B., Koehler, C. J., Strozynski, M., Treumann, A., Stein, R., Zimny-Arndt, U., Schmid, M. & Jungblut, P. R. (2013). High resolution quantitative proteomics of HeLa cells protein species using stable isotope labeling with amino acids in cell culture (SILAC), two-dimensional gel electrophoresis (2DE) and nano-liquid chromatograpohy coupled to an LTQ-OrbitrapMass spectrometer. *Molecular & Cellular Proteomics*, 12 (2): 529-538.
Tyanova, S., Temu, T. & Cox, J. (2016a). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols*, 11 (12): 2301.
Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M. & Cox, J. (2016b). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, 13: 731. doi: 10.1038/nmeth.3901
https://www.nature.com/articles/nmeth.3901#supplementary-information.
Van Soest, P. (1994). Nutritional ecology of the ruminant. Cornell Univ. Press, Ithaca, NY. *Nutritional ecology of the ruminant. 2nd ed. Cornell Univ. Press, Ithaca, NY.*: -.
Vanwonterghem, I., Jensen, P. D., Ho, D. P., Batstone, D. J. & Tyson, G. W. (2014). Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Current opinion in biotechnology*, 27: 55-64.
Vijayakumar, M., Park, J. H., Ki, K. S., Lim, D. H., Kim, S. B., Park, S. M., Jeong, H. Y., Park, B. Y. & Kim, T. I. (2017). The effect of lactation number, stage, length, and milking frequency on milk yield in Korean Holstein dairy cows using automatic milking system. *Asian-Australasian journal of animal sciences*, 30 (8): 1093.
Villares, A., Moreau, C., Bennati-Granier, C., Garajova, S., Foucat, L., Falourd, X., Saake, B., Berrin, J.-G. & Cathala, B. (2017). Lytic polysaccharide monooxygenases disrupt the cellulose fibers structure. *Scientific reports*, 7: 40262.
Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C.-A., Hyslop, J. J., Ross, D. W., Waterhouse, A., Watson, M. & Roehe, R. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics*, 16 (1): 839. doi: 10.1186/s12864-015-2032-0.
Wilkins, M. R., Sanchez, J.-C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F. & Williams, K. L. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and genetic engineering reviews*, 13 (1): 19-50.

Wilmes, P. & Bond, P. L. (2006). Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends in microbiology*, 14 (2): 92-97.
Wyman, C. & Yang, B. (2009). Cellulosic biomass could help meet California's transportation fuel needs. *California Agriculture*, 63 (4): 185-190.
Xiao, C. & Anderson, C. T. (2013). Roles of pectin in biomass yield and processing for biofuels. *Frontiers in plant science*, 4: 67.
Xue, D., Chen, H., Luo, X., Guan, J., He, Y. & Zhao, X. (2018). Microbial diversity in the rumen, reticulum, omasum, and abomasum of yak on a rapid fattening regime in an agro-pastoral transition zone. *Journal of Microbiology*, 56 (10): 734-743.
Yue, Z.-B., Li, W.-W. & Yu, H.-Q. (2013). Application of rumen microorganisms for anaerobic bioconversion of lignocellulosic biomass. *Bioresource technology*, 128: 738-744.
Yvon-Durocher, G., Allen, A. P., Bastviken, D., Conrad, R., Gudasz, C., St-Pierre, A., Thanh-Duc, N. & Del Giorgio, P. A. (2014). Methane fluxes show consistent temperature dependence across microbial to ecosystem scales. *Nature*, 507 (7493): 488.
Zougman, A., Selby, P. J. & Banks, R. E. (2014). Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics*, 14 (9): 1006-1000.

Appendix

Appendix A-1



Scatterplots with Pearson correlations for biological replicates for samples from cow identified in MQrun5. A) Scatterplot with Pearson correlation for all replicates from cow fed the MAP diet. The correlations varied from 0.751 and 0.924. B) Scatterplot with Pearson correlation for all replicates from cow fed the COS diet. The correlations varied from 0.417 and 0.812, as shown in **Figure 4.3**. C) Scatterplot with Pearson correlation for all replicates from cow fed the HPO diet. The correlations varied from 0.843 and 0.916. D) Scatterplot with Pearson correlation for all replicates from cow fed the CTL diet. The correlations varied from 0.584 and 0.893. Diet acronyms are as follows; CTL = basal control, no additional oil, HPO =

basal concentrate supplemented with hydrogenated palm oil, MAP = basal concentrate supplemented with marine algae powder, COS = basal concentrate containing wheat starch and supplemented with corn oil.

Appendix A-2

155 significant protein groups from ANOVA p-value=0.05 conducted with protein groups identified from MQrun5. 8 protein groups were annotated (InterPro annotation) to *Fibrobacter succinogenes* and are marked in red. 77 protein groups were annotated (InterPro annotation) to *Prevotella* strains and are marked in blue.

				Log 2 transformed LFQ intensities for samples from cow															
Organism	**Fasta header**	**Pepti-des**	**-Log ANOVA p-value**	**Cow1_P1_HPO**	**Cow1_P2_MAP**	**Cow1_P3_CTL**	**Cow1_P4_COS**	**Cow2_P1_MAP**	**Cow2_P2_COS**	**Cow2_P3_HPO**	**Cow2_P4_CTL**	**Cow3_P1_CTL**	**Cow3_P2_HPO**	**Cow3_P3_COS**	**Cow3_P4_MAP**	**Cow4_P1_COS**	**Cow4_P2_CTL**	**Cow4_P3_MAP**	**Cow4_P4_HPO**
Anaeromyces robustus v1.0	Anasp1|324654|	3	2,09	23,33	22,27	22,54		21,45		22,99	22,47	23,77	22,01		23,80	22,74	21,70	25,13	23,39
Bacteroidales bacterium UBA1173 [order]	UBA1173|contig_1337_92|	3	1,96				19,75		19,58	19,17	19,57		18,87	19,09		19,64			18,73
Bacteroidales bacterium UBA1173 [order]	UBA1173|contig_444_7|	8	2,80	21,24		19,42	21,35		20,74	19,75	19,48	20,26	21,71			20,61	20,17		20,11
Bacteroidales bacterium UBA1179 [order]	UBA1179|contig_13136_5|	2	1,60				19,08		18,91					19,42		19,40		19,03	19,49
Bacteroidales bacterium UBA1187 [order]	UBA1187|contig_10617_9|	1	1,79				21,27	17,00	21,63	17,43	18,63		17,27			18,67			
Bacteroidales bacterium UBA1192 [order]	UBA1192|contig_125499_15|	2	1,45				18,54		19,99					18,72					
Bacteroidales bacterium UBA1223 [order]	UBA1223|contig_35434_43|	3	2,26		18,68		18,83		20,43				19,73	21,08		22,49			
Bacteroidales bacterium UBA1253 [order]	UBA1253|contig_7143_12|	3	1,45						19,15							19,25			
Bacteroides sp. AR20	HUN1000|IE59DRAFT_03634|	3	1,36				20,48		20,13		19,20			20,57			20,03		
Bacteroides sp. Ga6A1	HUN1000|T544DRAFT_00176|	5	1,83						19,45	18,00				20,08		19,36			
Bacteroides sp. Ga6A2	HUN1000|T538DRAFT_01142|	6	1,57	19,07	18,97		19,88	19,44	19,43	19,06	19,12		19,56		19,99			21,32	20,82
Bacteroides sp. Ga6A2	HUN1000|T538DRAFT_01495|	11	2,33			20,84	20,45		22,99		19,69			21,03		21,43			20,66
Bacteroides vulgatus strain NLAE-zl-G202	HUN1000|Ga0066897_10053|	2	3,20				21,61		20,70					21,74	21,17	21,30			
Blautia schinkii DSM 10518	HUN1000|T506DRAFT_00583|	2	1,35	20,15			20,25						19,92	19,63		19,72			19,37

Butyrivibrio hungatei NK4A153	HUN1000|G628DRAFT_00989|	3	1,83						20,48					19,51				
Catabacter sp. UBA1260 [genus]	UBA1260|contig_407231_11|	3	2,41				17,84		22,58	17,29				20,44		22,66	18,84		16,47
Clostridiales	AN|HiSeq_06477480|	2	1,76											19,65		20,23			
Clostridiales	AS1a|HiSeq_01653190|	6	1,71		19,51		19,82		21,80				19,99	20,56					
Clostridiales bacterium NK3B98	HUN1000|G618DRAFT_00181|	7	1,92			19,17	19,18		18,85	19,22	20,15		19,47	20,31		19,92			
Clostridiales bacterium strain R-7	HUN1000|Ga0066890_102397|	5	1,51	19,20	19,17		20,68	19,86	19,94		19,44		19,52	22,99		20,47	20,14		
Desulfovibrio legallii strain KHC7	HUN1000|Ga0104380_101214|	5	3,74	19,15			20,06		19,84					20,19		22,57			
Dorea sp. AGR2135	HUN1000|G593DRAFT_01943|	13	3,12	22,45	21,55	22,73	20,81	22,23	20,97	22,21	21,82	22,41	21,23	20,67	22,53	20,74	22,46	22,76	22,04
Fibrobacter succinogens	Fibrobacter_succinogens|WP_012820155.1|	3	3,22	18,74	18,99		19,96	19,34	18,99	19,73			19,65	21,21		20,59			18,96
Fibrobacter succinogens	Fibrobacter_succinogens|WP_014546480.1|	14	2,31	20,96	21,03	22,01	24,87	21,80	23,48	20,97	20,92	20,45	20,29	25,82	18,70	22,96	21,31	21,76	
Fibrobacter succinogens	Fibrobacter_succinogens|WP_014546482.1|	3	2,19				19,99		19,49					20,56					
Fibrobacter succinogens	Fibrobacter_succinogens|WP_014547018.1|	2	1,82				19,17							20,40					
Fibrobacter succinogens	Fibrobacter_succinogens|WP_014547218.1|	3	1,65	19,60			20,31			20,11	20,38		20,15	21,27		19,88	20,22		
Fibrobacter succinogens	Fibrobacter_succinogens|WP_015732010.1|	3	2,02				20,23	19,27	19,33				19,23	20,46		19,60			
Fibrobacter succinogens	Fibrobacter_succinogens|WP_015732116.1|	1	1,35				19,47							19,11					
Fibrobacter succinogens	Fibrobacter_succinogens|WP_015732149.1|	4	4,34				20,55		19,47					20,04		20,34			
Gammaproteobacteria	UBA1249|contig_113521_4|	10	1,35		20,45		20,00		22,30				20,75	21,26		24,20	21,54	19,93	

bacterium UBA1249 [class]																		
Gammaproteobacteria bacterium UBA1249 [class]	UBA1249|contig_147162_33|	5	1,72				19,23		21,22		19,37			20,14		22,91	19,65	20,58	
Gammaproteobacteria bacterium UBA1249 [class]	UBA1249|contig_147162_38|	3	1,37						20,73							21,67			
Gammaproteobacteria bacterium UBA1249 [class]	UBA1249|contig_226955_1|	13	1,48				18,26		21,28	18,32			19,18	20,29		24,52	20,20	18,63	18,96
Gammaproteobacteria bacterium UBA1249 [class]	UBA1249|contig_226955_2|	4	1,51				17,93		20,71				19,39	19,41		21,88		19,27	
Gammaproteobacteria bacterium UBA1249 [class]	UBA1249|contig_284718_5|	4	1,64						20,02					19,95		21,23			
Gammaproteobacteria bacterium UBA1249 [class]	UBA1249|contig_39801_10|	3	1,92				19,30		19,70			18,85		19,49	18,82	21,69	19,80		
Gammaproteobacteria bacterium UBA1249 [class]	UBA1249|contig_553531_7|	9	1,45	18,17	19,70		17,88		23,71		16,80	17,85	19,21	20,74		24,02	19,59	18,29	18,50
Kandleria vitulina strain WCC7	HUN1000|Ga0104359_11751|	2	1,58				19,05							20,73		20,65			
Lachnospiraceae bacterium NC2008	HUN1000|T528DRAFT_01891|	9	1,93	22,23	20,56	22,08	20,26	21,24	19,90	21,13	21,06	21,68	21,57	20,78	21,93	20,71	22,32	21,81	22,43
Lachnospiraceae bacterium NE2001	HUN1000|IE09DRAFT_00779|	12	2,15	25,63	24,03	25,99	24,14	24,85	23,96	24,96	25,32	25,40	25,90	24,37	25,27	24,80	25,22	25,46	25,64
Lachnospiraceae bacterium NK4A136	HUN1000|G612DRAFT_00650|	10	1,37	23,31	23,45	23,94	22,40	22,90	21,57	22,89	22,67	23,48	22,59	22,40	23,66	22,53	22,50	22,90	23,24
Mageeibacillus sp. UBA1243 [genus]	UBA1243|contig_8436_15|	6	1,59	19,90		20,20						19,93	20,64						19,97
Methanobrevibacter millerae DSM 16643	HUN1000|IE19DRAFT_01533|	3	1,69	20,05		19,70				19,95	20,05		21,12	19,79		20,02	20,43		20,16
Morganella morganii strain NLAE-zl-C84	HUN1000|Ga0104395_11310|	5	1,69						18,89					19,63		20,86	18,50		
Orpinomyces sp. strain C1A	Orpsp1_1|1184370|	1	1,47	20,71		21,16		20,18		20,79	21,02	20,47	19,60		21,31			20,51	20,69
Piromyces sp. E2 v1.0	PirE2_1|67181|	3	2,06	21,93		23,13		21,83		21,92	21,49	21,73	20,60		21,76		20,05	22,85	21,33
Porphyromonadaceae bacterium UBA1250 [family]	UBA1250|contig_153986_6|	3	1,85	21,12	21,75	19,86	22,44	21,39	22,77	20,60	20,68	20,75	21,27	23,07	21,05	21,87	22,50	21,09	20,78

Prevotella brevis P6B11	HUN1000|T496DRAFT_00618|	10	1,64	22,38	22,44	22,02	23,82	22,95	23,67	22,84	23,40	22,92	21,76	23,00	22,42	23,53	22,27	23,12	22,42
Prevotella brevis P6B11	HUN1000|T496DRAFT_02494|	2	1,75				18,39						19,01	18,88		18,70			
Prevotella bryantii FB3001	HUN1000|Ga0066874_102219|	5	1,45						21,54					19,27		19,08			
Prevotella bryantii KHPX14	HUN1000|Ga0104364_106134|	7	1,76						21,77					20,46		20,79			
Prevotella bryantii KHPX14	HUN1000|Ga0104364_108182|	10	1,73				19,52	19,59	23,16		20,08			19,55		20,02			
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_00139|	2	2,02				18,57		18,17										
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_00259|	3	2,14				22,53		22,14	20,43	21,06		19,60	21,07		20,15			
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_00286|	3	2,65				20,50		20,24		20,60			20,65		20,30			
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_00628|	4	2,58				19,96		19,84					21,58					
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_00637|	2	2,70		18,37		21,73		21,73	20,13	20,69	18,87	21,07	22,07		21,53			20,35
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_00652|	5	1,46				22,22	20,58	21,49	20,75	21,67	20,59	20,19	21,85		21,37	20,55		20,58
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_00708|	5	1,66		19,16		22,56		21,74	21,07	20,98			20,45		20,85			
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_00750|	3	1,68				21,56	19,30	19,81		19,79			20,97					
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_00752|	4	1,97		21,52	19,56	23,43		23,47	20,70	20,55		21,04	23,78		21,78	20,26		20,81
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_01115|	8	1,52	20,76	21,82	19,83	21,93	21,57	22,60	21,17	21,76	20,24	20,93	22,39	20,18	22,15	21,65	20,59	20,80
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_01162|	13	3,14	22,59	22,89	21,74	23,52	23,07	24,30	22,67	23,06	21,78	22,41	24,06	22,78	23,44	22,00	22,24	22,27
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_01172|	4	1,31	20,07	20,60	20,19	22,10	20,40	22,05	21,35	21,70	19,02	20,06	21,46	19,46	21,04	20,33	20,36	20,04
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_01268|	5	2,73				21,18		19,99					20,66					
Prevotella ruminicola Ga6B6	HUN1000|T500DRAFT_01484|	3	1,48				22,18		21,36	21,37	21,46		21,39	22,69		22,08		21,00	

Prevotella ruminicola KHT3	HUN1000|Ga0104360_101141|	2	1,49				19,77							19,18				
Prevotella ruminicola KHT3	HUN1000|Ga0104360_10121|	8	2,44	20,35	21,15	20,76	22,18	20,57	21,97	20,65	21,86	20,24	20,99	22,03	20,49	21,61	20,45	20,20	20,09
Prevotella ruminicola KHT3	HUN1000|Ga0104360_10226|	2	1,99				21,17		19,23										
Prevotella ruminicola KHT3	HUN1000|Ga0104360_10452|	3	1,65				21,76		20,75	20,33						20,37	20,48		
Prevotella ruminicola KHT3	HUN1000|Ga0104360_107113|	4	1,34				20,00		19,87	19,19	19,69			20,29					
Prevotella ruminicola KHT3	HUN1000|Ga0104360_108102|	5	1,34				21,50	20,57	20,19	19,84	19,49	18,97	20,03	21,20	19,93	20,57		19,44	
Prevotella ruminicola KHT3	HUN1000|Ga0104360_10868|	4	3,08				20,53		18,63	18,93	19,08			22,12		21,16			
Prevotella ruminicola KHT3	HUN1000|Ga0104360_11032|	3	1,74				19,83		19,95										
Prevotella ruminicola KHT3	HUN1000|Ga0104360_1104|	9	1,65				18,81		18,28				19,58	20,12		20,07		19,52	
Prevotella ruminicola KHT3	HUN1000|Ga0104360_11611|	3	1,30				19,29		19,51		19,32			20,08					
Prevotella ruminicola KHT3	HUN1000|Ga0104360_1169|	10	2,27	20,08	20,94	19,78	20,99	20,79	22,36	20,24	19,88	19,85	19,58	21,60	19,99	20,71	20,58	20,00	19,97
Prevotella ruminicola KHT3	HUN1000|Ga0104360_12014|	4	3,25				20,87		21,38		20,25			21,38		20,69			
Prevotella ruminicola KHT3	HUN1000|Ga0104360_1225|	4	1,34	19,24			21,07	19,78	21,27					19,84					
Prevotella ruminicola strain BPI-162	HUN1000|Ga0070636_102222|	2	1,90						18,94					19,62		20,06			
Prevotella ruminicola strain BPI-162	HUN1000|Ga0070636_104165|	16	1,85				21,73	18,76	22,33	21,09	22,05		20,38	22,70		21,70	19,09		19,62
Prevotella ruminicola strain BPI-162	HUN1000|Ga0070636_105131|	4	1,44						20,21	19,83	20,16		20,19	21,51		21,25			
Prevotella ruminicola strain BPI-162	HUN1000|Ga0070636_10923|	6	1,77	18,97	20,20	18,08	21,94	19,68	21,65	20,42	20,16	20,22	21,13	22,12	19,08	21,60	20,74	20,65	20,72
Prevotella sp. AGR2160	HUN1000|G604DRAFT_00411|	2	2,24				19,89	19,66	21,13					20,57					
Prevotella sp. AGR2160	HUN1000|G604DRAFT_00879|	4	2,35				20,57		20,86		19,75		20,48	22,09		20,75	19,86		

Prevotella sp. AGR2160	HUN1000|G604DRAFT_00908|	4	1,50										20,86		20,20			
Prevotella sp. AGR2160	HUN1000|G604DRAFT_00911|	13	1,79		20,61	17,42	19,69	19,96	23,22	18,71	20,23		19,46	23,37		22,00			20,15
Prevotella sp. AGR2160	HUN1000|G604DRAFT_01101|	6	1,36						20,74					21,37				
Prevotella sp. AGR2160	HUN1000|G604DRAFT_01342|	8	1,35						22,71					21,56				
Prevotella sp. AGR2160	HUN1000|G604DRAFT_01943|	14	2,37						24,40					23,55		21,71			
Prevotella sp. AGR2160	HUN1000|G604DRAFT_01995|	6	1,97						21,54							20,19			
Prevotella sp. BP1-145	HUN1000|Ga0066893_11625|	7	1,96	20,85	22,26	20,94	22,63	21,43	22,65	21,14	21,60	20,41	22,22	22,75	21,02	22,04	21,50	21,85	20,65
Prevotella sp. BP1-145	HUN1000|Ga0066893_11629|	5	1,94	18,31		17,66	18,62	17,78	20,18	18,67	19,10		18,61	20,68	17,67	19,91	18,68		18,85
Prevotella sp. BP1-148	HUN1000|Ga0066894_102125|	6	1,36	20,80	21,33		23,16	22,04	22,69	21,78	20,92	20,75	20,59	22,50	21,15	21,02	21,05	20,91	21,01
Prevotella sp. BP1-148	HUN1000|Ga0066894_102165|	8	1,78				20,04		20,00	19,45	20,00		19,87	21,35		20,65			19,46
Prevotella sp. BP1-148	HUN1000|Ga0066894_104163|	10	1,35			19,25	20,45		20,79	19,90	19,82		20,29	21,67	18,34	20,02		19,57	19,74
Prevotella sp. BP1-148	HUN1000|Ga0066894_10513|	12	2,11	21,83	21,99	21,43	22,38	21,87	22,84	21,72	22,14	21,59	21,67	22,31	21,47	22,33	22,06	22,22	21,53
Prevotella sp. BP1-148	HUN1000|Ga0066894_10658|	2	1,47				19,10		18,21					19,64					
Prevotella sp. BP1-148	HUN1000|Ga0066894_10941|	5	1,48	19,98			22,35	20,60	22,35	20,25	22,09		19,65	21,26		20,18			19,75
Prevotella sp. BP1-148	HUN1000|Ga0066894_11453|	2	1,35				21,03	20,68	20,97	20,36		20,30		21,57		20,48		20,21	
Prevotella sp. kh1p2	HUN1000|Ga0059114_11548|	3	2,39				20,66		21,78	18,78	19,50		19,14	21,77		20,86	19,74	19,10	19,59
Prevotella sp. KH2C16	HUN1000|Ga0104366_11338|	2	1,72				19,66		21,58			19,68		19,53					
Prevotella sp. KH2C16	HUN1000|Ga0104366_13411|	13	2,11					20,30	25,04					21,77		21,77			
Prevotella sp. NE3005	HUN1000|Ga0066875_10121|	4	1,40		20,41		22,00		22,00	20,17	20,76		20,36	21,62		20,45	20,96		20,57

Prevotella sp. NE3005	HUN1000|Ga0066875_10155|	21	2,86	22,17	22,71	21,57	26,04	24,20	26,48	23,06	23,63	20,55	22,60	26,70	22,67	24,90	24,28	23,23	22,89
Prevotella sp. NE3005	HUN1000|Ga0066875_102110|	13	1,60	22,60	21,80	22,87	23,91	22,68	23,48	23,37	23,65	22,38	22,49	23,65	22,25	23,64	22,47	22,55	21,87
Prevotella sp. NE3005	HUN1000|Ga0066875_102224|	2	1,93				21,81							22,04					
Prevotella sp. NE3005	HUN1000|Ga0066875_103174|	2	1,58			19,63	21,95		20,42										
Prevotella sp. NE3005	HUN1000|Ga0066875_104134|	1	1,71						19,16					19,70					
Prevotella sp. NE3005	HUN1000|Ga0066875_105284|	4	1,62						21,53	20,47				20,83		20,37			
Prevotella sp. NE3005	HUN1000|Ga0066875_10882|	9	1,66	21,06	21,04	19,90	23,18	20,95	23,35	22,38	23,36	19,56	22,41	23,61	20,12	22,94	20,89	22,04	22,07
Prevotella sp. NE3005	HUN1000|Ga0066875_11142|	14	2,13	22,47	22,54	21,70	23,12	22,22	23,16	22,42	22,69	22,03	22,22	23,46	21,66	22,60	22,21	22,15	21,84
Prevotella sp. NE3005	HUN1000|Ga0066875_11227|	6	1,87	20,94	21,08	20,94	21,80	21,25	22,88	21,42	21,80	20,73	20,84	21,93	21,21	21,59	21,44	21,08	20,57
Prevotella sp. NE3005	HUN1000|Ga0066875_11230|	3	2,40				20,58		19,57							19,73			
Prevotella sp. P6B4	HUN1000|T491DRAFT_00392|	12	1,43	22,12	22,05	20,70	23,20	22,53	23,89	21,96	22,04	21,72	22,28	23,61	21,93	22,57	23,34	21,31	21,66
Prevotella sp. P6B4	HUN1000|T491DRAFT_00572|	4	1,42	19,34	19,96		19,76	19,56	20,60	18,67	19,45		19,73	20,35		19,67			
Prevotella sp. P6B4	HUN1000|T491DRAFT_01129|	5	1,45				20,56		20,71	19,54	20,72		19,90	21,10					19,09
Prevotella sp. P6B4	HUN1000|T491DRAFT_02725|	4	1,56	20,72	20,67	20,74	21,62	20,54	21,47	20,99	21,34	20,44	20,08	21,24	20,45	20,79	20,72	20,56	20,19
Prevotella sp. P6B4	HUN1000|T491DRAFT_02932|	2	1,44			17,67	21,80	17,35	20,98	20,85	22,26	17,72	20,16	22,11		20,28	20,54	18,01	
Prevotella sp. TC2-28	HUN1000|Ga0066886_102266|	3	3,52				19,88		20,44	19,31			19,47	20,53		19,59			19,16
Prevotella sp. TC2-28	HUN1000|Ga0066886_10667|	5	1,46		18,82		18,80	19,16	20,60	19,03	19,05	19,29		19,46	19,00	20,56	19,12	18,70	
Prevotella sp. TC2-28	HUN1000|Ga0066886_111102|	4	6,17				20,74		21,05					21,08		20,47			
Prevotella sp. TF2-5	HUN1000|Ga0066888_10676|	7	1,69	21,74	21,75	21,88	22,99	21,68	23,61	21,78	22,08	21,54	22,19	22,79		22,44	22,20		21,53

Prevotella sp. TF2-5	HUN1000|Ga0066888_10744|	5	1,96		20,03			20,52	21,86					19,93		21,70			
Prevotellaceae bacterium KH2P17	HUN1000|Ga0066882_102102|	9	1,51						22,77					21,76		21,46		20,82	
Prevotellaceae bacterium KH2P17	HUN1000|Ga0066882_11262|	9	1,89	22,34	20,40	21,38	21,88	20,59	24,44	21,47	21,55	20,49	22,22	24,30	21,34	22,99	22,24	22,30	21,55
Prevotellaceae bacterium KH2P17	HUN1000|Ga0066882_11544|	5	1,52		21,49		22,29		22,36	20,96	20,71		20,74	21,02		21,73			20,72
Prevotellaceae bacterium MN60	HUN1000|Ga0066892_101426|	3	3,04				19,75		19,69					19,88					
Prevotellaceae bacterium MN60	HUN1000|Ga0066892_10311|	2	2,21		18,56		19,15		20,25					19,93					
Prevotellaceae bacterium MN60	HUN1000|Ga0066892_104110|	8	1,86				20,11		20,10	19,50	19,53		19,53	20,15		19,71	19,54		
Prevotellaceae bacterium MN60	HUN1000|Ga0066892_104136|	4	1,31	19,99	20,58	20,10	22,45	21,53	22,13	21,87	22,32	20,63	21,51	22,27	20,47	22,41	21,40	21,14	20,93
Prevotellaceae bacterium MN60	HUN1000|Ga0066892_10448|	2	1,68				20,59							19,76					
Prevotellaceae bacterium MN60	HUN1000|Ga0066892_11241|	8	1,57	21,25	22,18	20,75	22,59	22,38	22,75	21,58	21,62	21,15	20,28	22,15	21,44	22,18	22,02	21,56	21,82
Ruminobacter sp. RM87	HUN1000|T489DRAFT_02180|	8	1,42				19,38		20,32			18,56	21,67	19,45	18,40	22,17			
Ruminococcus flavefaciens ND2009	HUN1000|T488DRAFT_02463|	7	2,31	22,74	21,92	22,89	20,47	22,27	20,85	22,55	21,74	23,42	21,35	19,57	23,06		21,15	22,16	22,13
Ruminococcus flavefaciens strain YRD2003	HUN1000|IE38DRAFT_00463|	2	2,18	20,13			20,24		22,58	19,98		20,22	20,03	20,38		19,67			
Ruminococcus flavefaciens strain YRD2003	HUN1000|IE38DRAFT_01026|	7	1,37	20,57	21,06		22,33	21,01	23,12	20,57	20,42	20,87	20,83	23,47	20,34	22,22	21,45		20,70
Sarcina sp. DSM 11001	HUN1000|Ga0059085_105103|	15	1,63	24,68	24,35	25,96	22,26	25,16	22,01	24,73	24,63	24,78	24,31	23,12	24,82	24,52	23,29	24,87	25,06
Streptococcus equinus JB1	HUN1000|IE46DRAFT_00424|	5	2,41	23,23	20,49	23,66	18,88	19,53		21,47	19,59	23,15	21,51	18,23	20,57	18,49	19,86	19,88	21,71
Succinatimonas sp. UBA1264 [genus]	UBA1264|contig_382805_5|	7	2,10				18,64		23,65	18,33	18,18			19,97		19,63			
Succinatimonas sp. UBA1264 [genus]	UBA1264|contig_528045_4|	2	1,86						22,64					20,08		21,94	19,48		
Succiniclasticum ruminis DSM11005	HUN1000|Ga0066895_102236|	7	1,71	21,95	20,58	20,84	23,19	20,73	23,55	22,54	23,75	21,46	23,65	24,09	21,13	23,56	22,13	22,75	22,93

Succinivibrio dextrinosolvens	HUN1000|T508DRAFT_01064|	3	2,19				17,76		24,08					21,68		20,93			
Succinivibrio dextrinosolvens	HUN1000|T508DRAFT_02079|	2	2,37		20,29				24,37					21,74		21,47			
Succinivibrio dextrinosolvens 22B	HUN1000|Ga0066883_100358|	7	2,89		18,27		18,73		23,45					19,96		22,73			
Succinivibrio dextrinosolvens strain ACV-10	HUN1000|IE82DRAFT_00591|	4	1,43						22,96					20,16					
Succinivibrio dextrinosolvens strain ACV-10	HUN1000|IE82DRAFT_00934|	5	1,52						23,38					20,46					
Succinivibrionaceae bacterium UBA1220 [family]	UBA1220|contig_191566_5|	4	1,78				20,13						20,97	20,85		22,39			
Treponema bryantii NK4A124	HUN1000|G626DRAFT_00568|	1	1,56				19,73		19,38										
Treponema bryantii strain B25	HUN1000|Ga0070255_11226|	4	1,95		20,76		20,87		20,68					21,91					

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway