2017

Identification of Candidate Cellulose Utilizing Bacteria from the Rumen of Beef Cattle, Using Bacterial Community Profiling and Metagenomics

Lee Opdahl

South Dakota State University

IDENTIFICATION OF CANDIDATE CELLULOSE UTILIZING BACTERIA FROM THE RUMEN OF BEEF CATTLE, USING BACTERIAL COMMUNITY PROFILING AND METAGENOMICS

BY

LEE OPDAHL

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Animal Science

South Dakota State University

2017

IDENTIFICATION OF CANDIDATE CELLULOSE UTILIZING BACTERIA FROM THE RUMEN OF BEEF CATTLE, USING BACTERIAL COMMUNITY PROFILING AND METAGENOMICS

LEE OPDAHL

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Animal Science degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Benoit St-Pierre, Ph.D.
Thesis Advisor Date

Michael Gonda, Ph.D.
Thesis Advisor Date

Joseph Cassady, Ph.D.
Head, Department of Animal Science Date

Dean, Graduate School Date

ACKNOWLEDGEMENTS

First, I would like to extend my gratitude to both of my advisors, Dr. Benoit St-Pierre and Dr. Michael Gonda. Dr. St-Pierre, thank you for always taking the time to explain processes and concepts both related and unrelated to my research. Your comprehensive knowledge and experience in the field of microbiology, as well as other fields, has been instrumental to my growth as a scientist, and will not be forgotten. Dr. Gonda, thank you for always taking the time to answer my questions and for offering your unending support. The advice and wisdom you have shared with me throughout my master's program was a key component to my success as a graduate student.

To my fellow graduate students, thank you for your support and assistance with my research throughout my time here at SDSU. To Ralco Agriculture, thank you for your financial support and your belief in me as a young scientist. I look forward to applying all that I have learned to continue the success and growth of the company.

Last, but certainly not least, I would like to thank my family. Mom, dad, Alyssa, Todd, Tim, Parker, Ryder, Cole, and Carver, thank you for always offering relief to my busy schedule, and for all of your support and encouragement. Your belief in me goes farther than you will ever realize.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

IDENTIFICATION OF CANDIDATE CELLULOSE UTILIZING BACTERIA FROM THE RUMEN OF BEEF CATTLE, USING BACTERIAL COMMUNITY PROFILING AND METAGENOMICS

LEE OPDAHL

2017

The ruminal microbiome allows ruminant animals to convert cellulosic biomass into food products. A majority of ruminal microorganisms remain uncharacterized due, in part, to the complexity of ruminal microbial communities. In order to gain further insight, selection-based batch culturing from bovine rumen fluid, in combination with metagenomics, was used to identify and characterize previously uncharacterized rumen bacteria capable of metabolizing cellulose, which was supplemented as a purified substrate. 16S rRNA-based population analysis was used to identify rumen bacteria enriched within 14 days of culturing. As a result of 4 independent experiments, seven different candidate cellulose-utilizing species-level operational taxonomic units (OTUs) were identified. Six of the enriched OTUs showed increased levels ranging between 46 and 445-fold compared to their respective rumen inocula, representing 14.1% to 41.3% of reads in samples supplemented with cellulose. One OTU corresponded to a known species (*Ruminococcus flavefaciens*), four OTUs were predicted to be uncultured species of known genera (*Ethanoligenens* sp., two *Prevotella* sp., and *Rummeliibacillus* sp.), and two were assigned to the family *Ruminococcaceae*. One enriched culture consisting of an uncultured *Rummeliibacillus* and *Prevotella* was used for metagenome analysis. Analysis

revealed genes with predicted cellulolytic capabilities in the *Rummeliibacillus*-related organism (cellulase, endoglucanase, and beta-glucanase) and in the *Prevotella*-related organism (cellulase). Additionally, genes predicted to function in cellulose binding, as well as proteases and glutamate synthases needed for amino acid acquisition, were also found in both OTUs. The identification and characterization of novel cellulolytic species of ruminal bacteria will contribute to a better understanding of ruminal cellulose metabolism.

Chapter 1

LITERATURE REVIEW

Global Challenge: Ensuring Food Security for a Growing Human Population

The global human population reached 7 billion in the year 2011, and is expected to exceed 9 billion before 2045 (Dawson & Johnson, 2014). Concerns arise when considering necessary resources to meet the demands of this growing population, namely food production. Prediction models show that global food demand is projected to increase by 74% by the year 2050 (Valin *et al.*, 2014). While many solutions to this problem are being investigated, simply increasing the sheer number of livestock or creating more cropland is unrealistic because of limited land availability and, it would not overcome the large deficit in food production (Bodirsky *et al.*, 2015). Achieving greater efficiency in food production is a more plausible solution to meet a growing global food demand. In countries, such as the United States, where livestock is a major commodity and contributor to food production, improved efficiency in livestock production would have significant economic benefits, particularly in regions such as South Dakota where the research described in this thesis was done.

The cost of feed is arguably the most limiting factor in animal production (Byerly, 1967). In order to overcome the high costs of traditional feed, alternative feedstuffs can be used; however, decreased dietary value of some alternative feeds may not provide the energy necessary for optimal animal production. Increasing feed efficiency in animals can not only increase food production, but also increase the economic margin for the producer (Thornton, 2010). Nutritional needs of livestock have been researched for decades and have progressively been refined, and a continuation of research aiming to improve feed efficiency is crucial to increase food production. However, feed efficiency in cattle is variable and can be affected by many extrinsic and intrinsic factors. Extrinsic

factors that impact variation in efficient use of feed by cattle include feed composition and quality, growth promotants, and environmental conditions, and animal-dependent intrinsic factors include host genetics, microbial milieu, and body composition (Van Soest, 1994; Bruns *et al.*, 2005; Barendse *et al.*, 2007; Kerr *et al.*, 2015). Of the factors that play an integral role in the overall efficiency of the animal, extrinsic factors that are more easily controlled, such as management and feed composition and quality, have long been studied in order to obtain optimal feed efficiency in comparison to intrinsic factors that impact efficiency.

The ability to breakdown plant biomass into usable energy is the quintessential role of the ruminant forestomach, distinguishing it from the gastrointestinal tract of non-ruminants. Thus, a necessary approach to further improving feed efficiency in cattle is to identify gaps in our current knowledge relating to feed breakdown. Cellulose, the most abundant polysaccharide on earth, and a major component of ruminant diets, can be broken down into metabolites (volatile fatty acids) that can be absorbed by the host to use as a source of energy (Naas *et al.*, 2014). The efficiency of a ruminant's ability to convert plant biomass into meat, milk or wool products is a growing area of interest, and further knowledge of this process could be useful in developing strategies to meet a growing food demand. Furthermore, cellulose is not the only substrate metabolized by ruminal microorganisms. Substrates such as starch, fat, and protein are also metabolized by ruminal microbiota into utilizable by-products for host absorption. In fact, the majority of the host's daily protein requirements are fulfilled through absorbed microbial proteins resulting from bacterial cell lysis (Yeoman & White, 2014).

The Rumen and its Microbial Symbionts

Rumen Function

One of the major differences between ruminants and non-ruminants are the anatomical differences of their gastrointestinal tracts. Ruminants possess a compartmentalized stomach consisting of four chambers: rumen, reticulum, omasum, and abomasum. These four "sac-like" structures vary in size and their epithelial structure, and differ for various activities such as absorption of nutrients (Van Soest, 1994). The rumen is where the majority of feed digestion takes place. While some feed components can be metabolized by the host (e.g. protein and fat), metabolism of fibrous plant material cannot be carried out by the host, as the enzymes required for this function have yet to be found in an animal genome, but are expressed by microbial inhabitants found in their gastrointestinal tract (Yue *et al.*, 2013). With the rumen being a mostly anoxic environment, rumen microorganisms rely on fermentative processes to fulfill their energetic needs (Reddy & Peck, 1978). Volatile fatty acids (VFAs) are by-products of fermentation that are used to fulfill approximately 60-80% of the daily metabolizable energy intake by ruminant animals (Weimer *et al.*, 2010; McSweeney & Mackie, 2012). Among the different VFAs that can be produced by rumen microbes, acetate, propionate, and butyrate are the most abundant and have distinct metabolic fates. In terms of development, butyrate has been shown to increase growth of ruminal papillae by increasing blood flow to these tissues during absorption and metabolism. Gene expression was also shown to be affected by increased butyrate concentrations in the rumen, showing an upregulation in genes related to metabolism and growth (Wang *et al.*, 2016).

Establishment of the Rumen Microbiome

A newborn ruminant displays anatomical and functional similarities to that of non-ruminants. At this time, various enzymes found in adult ruminants are not yet expressed and, as a result, diets must be limited, typically consisting of milk or milk replacers, providing glucose and lactose as the primary sources of energy for the young calf (Van Soest, 1994). The liquid ingested by the infant calf flows down the esophagus, and is directed to the abomasum via the esophageal groove. Because the esophageal groove bypasses the rumen, it inhibits ingesta from being fermented. Following a regimental diet consisting of milk, calves are eventually introduced to solid-type feeds. The bolus formed upon ingesting solid feed enters directly into the rumen, which marks the beginning of considerable rumen development by the calf (Muya *et al.*, 2015). Rumen development in calves is therefore dependent upon two primary factors: introduction of solid feed in the diet, and microbial colonization of the rumen.

Members of the three domains of life, bacteria, eukarya (protozoa and fungi), and archaea, are all found in high abundance in the rumen. Of the three domains, bacteria are found at the highest cell density with approximately 10^{11} cells per milliliter, indicating their importance in this complex ecosystem (Weimer *et al.*, 2009). It has been shown in dairy calves that ruminal functions, likely correlating to microbial activity, can occur in the first 48 hours of life, indicating that a precursory colonization of microbes occurs before the introduction of fibrous feeds. Inoculation is likely from microbes found in the milk, mammary teat surface, and the environment in which the calf is held (Rey *et al.*, 2012). In a study by Rey *et al*. (2014), the rumen microbiome of a dairy calf in the first two days of life was found to be dominated by members of the phylum *Proteobacteria*. The dominance of *Proteobacteria* then rapidly transitions to members of the phylum

Bacteroidetes. Species from both *Bacteroidetes* and *Proteobacteria* have been shown to produce VFAs, supporting their importance at the early stages of life in a newborn calf (Van Soest, 1994). This transition observed in the rumen microbiome of dairy calves correlates to changes in calf diet.

As cattle develop and are introduced to new feedstuffs, the rumen microbiome also changes. Li *et al.* (2011) measured the relative distribution of bacterial phyla in the rumen up to 12 months of age. Their findings showed the emergence of the phylum *Firmicutes* at the age of one year. It has generally been observed that *Bacteroidetes* and *Firmicutes* are the most prevalent bacterial phyla in the rumen of adult cattle. While fluctuations in abundance can occur, these two groups typically dominate regardless of environmental changes (McSweeney & Mackie, 2012).

Microbial Ecology of the Rumen

Anaerobic microorganisms tend to be specialized in the metabolic function(s) they provide towards maintaining their community structure. Limitations in metabolic function are primarily due to relatively small individual genome size, limiting the expression of proteins needed to supply all demands of the cell. In microbial anaerobic communities, metabolic breakdown of substrates can be divided into four major groups: hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Thauer *et al.*, 2008; Angelidaki *et al.*, 2011). Organisms with hydrolytic capabilities break down large polymers such as polysaccharides, proteins, and lipids into their monomeric subunits, which can then be metabolized into organic acids through acidogenesis. Organic acids can be further utilized by acetogenic microorganisms to create compounds such as acetate, formate, carbon dioxide, and hydrogen gas. Lastly, terminal products such as

hydrogen gas and carbon dioxide, are used to create methane, a fully reduced end-product, by specialized groups of archaea in a process known as methanogenesis (Thauer *et al.*, 2008; Ferry, 2010). In this "assembly line" type metabolism, codependence is seen at each level where microorganisms performing downstream catabolism are highly dependent on the metabolic activities of those microbes that initially breakdown feedstuffs, while microbes performing upstream catabolism rely on microorganisms to breakdown and utilize their end products to provide favorable forward reactions.

Individually, anaerobic microbial species tend to have specialized metabolic functions, and therefore often depend on complementary community processes to fulfill their metabolic needs (Johnson *et al.*, 2012). Because the survival of ruminal microorganisms is highly dependent on the metabolic characteristics of other community members, the complex assemblage of thousands of different species represents an intricate, highly sophisticated network that is able to resist potential changes within the rumen microenvironment. The resiliency and host specificity of ruminal microbes is a remarkable characteristic seen in studies where a new species of bacteria is inoculated into a cannulated animal whose ruminal contents have been emptied. Within several hours, the bacterial strains from the inoculated culture were cleared to undetectable levels (Varel *et al.*, 1995; Weimer *et al.*, 2010).

Investigative Tools for Rumen Microbial Communities

Culture-Dependent Techniques

Complex anaerobic microbial communities have been a subject of interest due to their direct and indirect influence on their surrounding environment (Ze *et al.*, 2013). For

many years, cultivation-based methods were the only tools available for characterization of these communities. Culture-dependent methods allow researchers to gain a deeper understanding of the metabolism, function, and growth requirements of specific microbes, permitting further investigation of individual microbial genomes, leading to a better understanding of uncharacterized genes, and ultimately giving the ability to manipulate microbial DNA (Pham & Kim, 2012). While allowing the characterization of bacterial species, culture-dependent methods are limited in revealing community compositions of complex microbial ecosystems (Margesin & Miteva, 2011; Rosling *et al.*, 2011). Due to our lack of understanding, only a fraction of the bacteria, archaea, fungi, and protists have been identified by means of culture-dependent methods (Epstein, 2013). While still largely a mystery, proposed explanations regarding the difficulty in culturing most anaerobic organisms include the lack of understanding of: 1) optimal environmental and physiochemical conditions for individual organisms, which are necessary for the development of media composition, and 2) symbiotic relationships essential to survival (Puspita *et al.*, 2012).

Culture-Independent Techniques

The development of culture-independent techniques took advantage of DNA sequencing technologies, allowing for taxonomic classification of microorganisms based on genomic comparisons. Rumen microbial diversity estimates increased from 20 predominant species per ruminant host using cultivation techniques to over 300 species as a result of using clone libraries and Sanger-based sequencing methods (Krause & Russell, 1996; Edwards *et al.*, 2004). Approximately 40 years after the advent of Sanger sequencing came a revolutionary breakthrough in the world of molecular biology, with

the creation of high throughput or next-generation sequencing (NGS) (Sogin *et al.*, 2006). NGS, in conjunction with bioinformatics tools, has equipped scientists with a new and improved way to study microbial ecology, revealing the existence of greater than expected diversity in ruminal microbial communities. By providing a more in-depth analysis of the rumen microbiome, NGS estimates have increased to several thousands of microbial species present per bovine host species (Fouts *et al.*, 2012; Jami & Mizrahi, 2012; Creevey *et al.*, 2014). The ability to assess the taxonomic composition of the rumen microbiome can be used to monitor microbial population variation throughout the life cycle of an animal, among different host animals, or as a result of varying treatment parameters (Henderson *et al.*, 2015).

Metagenomics

Metagenomics is an approach that has gained increased popularity as a result of advancements made in NGS technologies. This –omics approach allows researchers to conduct a sequence and functional-based analysis of microbial genomes collected from an environmental sample (Morgavi *et al.*, 2013). Shotgun short read sequencing of microbial samples, followed by assembly of the reads into contigs, has allowed researchers to determine partial genomes of novel microorganisms in the environment sampled. Comparative analysis not only reveals taxonomic information, but more importantly, open reading frames to functionally annotate genomic sequences can identify potential metabolic properties of the microorganisms in the sample (Handelsman *et al.*, 1998). Ferrer *et al*. (2005) pioneered this approach when they identified and characterized 12 esterases, 9 endoglucanases, and 1 cyclodextrinase from a metagenomic library obtained from the rumen of a dairy cow. In addition to uncovering enzymes

involved in the hydrolysis of plant cell wall polysaccharides, opportunities also exist to use metagenomics for the discovery of genes involved in biotransformation of compounds of interest, such as conjugated linoleic acids and antioxidants, or the detoxification of toxins derived from plants or fungi (Morgavi *et al.*, 2013).

Improvements in bioinformatics has greatly increased our ability to sort through large quantities of sequence data, notably through the comparison of patterns in nucleotide sequence that are unique to individual microorganisms (Zhou *et al.*, 2008; Patil *et al.*, 2011; Droge & McHardy, 2012; Albertsen *et al.*, 2013). As a result of continual enhancement of informatics technology, functionally-annotated genes obtained from metagenomics studies can be used to model rumen microbial metabolism (Raes & Bork, 2008; Wang *et al.*, 2013; Jose *et al.*, 2017). However, while metagenomics provide valuable insights into the functionality of microorganisms present in an environment of interest, linking function to phylogeny can be difficult due to insufficient database availability of homologous sequences that have been assigned biochemical or biological roles using experimental models (Denman & McSweeney, 2015).

Single-Gene Markers

Metagenomics would be more comprehensive if the entire genome of all rumen-dwelling microorganisms could be sequenced. However, due to technology limitations and cost restraints of this approach, universal single-gene markers provide a more complete phylogenetic assessment (Morgavi *et al.*, 2013). First described by Lane *et al*. (1985), the 16S rRNA is the most widely used phylogenetic marker for identifying and classifying prokaryotes in any given environment. As of 2014, nearly 3 million bacterial and archaeal 16S rRNA sequences had been archived in the Ribosomal Database Project

(RDP) library (Cole *et al.*, 2014). This rapid accumulation of 16S rRNA gene sequences can be attributed to the growing use of NGS, and the number of sequences archived is expected to further increase as the number of metagenomics studies of microbial communities continues.

The 16S rRNA gene possesses 4 primary qualities making it the preferred phylogenetic marker: 1) it is an essential gene for protein synthesis and thus the function of the gene has not changed over time, 2) it is thought to be present in all prokaryotes, 3) the presence of conserved regions that tend to show limited variations in sequence across vast phylogenetic distances which can be used for primer targeting, and 4) the presence of hypervariable regions allowing differentiation amongst different taxa (Patel, 2001). Nine distinct hypervariable regions have been identified within this prokaryotic gene, each varying in length and sequence, and they are key to the classification of either bacteria or archaea (Vinje *et al.*, 2014). While sequencing of the full length gene provides the most accurate phylogenetic assignment, 16S rRNA-based community analyses are currently done in most cases using partial sequences of the 16S rRNA gene. Limitations in NGS technology for sequence length as well as the high cost associated with Sanger sequencing of clone libraries limit our ability to sequence full length genes (Kim *et al.*, 2011b). While using partial 16S rRNA gene sequences may be the most comprehensive approach to determine the phylogeny of members of a poorly characterized microbial community, shorter fragments of DNA may not be assigned to taxonomic groups as accurately as longer fragments that are greater than 1kb in length (Patil *et al.*, 2011).

In diversity studies, using genes that are ubiquitous to all prokaryotes, such as the 16S rRNA gene or the *rpoB* gene, is essential to analyzing phylogenetic diversity in a

given environment. However, the main limitation of the 16S rRNA gene is that phylogeny or taxonomic assignment cannot be used to infer function unless it is matched to a known species (Case *et al.*, 2007; von Mering *et al.*, 2007). For studies aiming to determine bacterial functionality, the use of single-gene markers such as *amoA* (NH3 oxidation)*, pmoA* (CH4 oxidation)*, nirS* (N reduction)*, nirK* (N reduction)*, nosZ* (N2O reduction)*, pufM* (anoxygenic photosynthesis), and [NiFe] hydrogenase genes, that are specific to particular functional groups of bacteria represent better targets (Achenbach *et al.*, 2001; Nicolaisen & Ramsing, 2002; Throback *et al.*, 2004; Hendrickx *et al.*, 2006).

The goal of single-gene marker use is to gain a deeper understanding of microbial communities by resolving their composition and diversity. Thus, two main strategies are currently used to analyze genomic sequences derived from single-gene marker diversity studies: taxonomic dependent and taxonomic independent binning. Taxonomic dependent methods rely on homology between the experimentally obtained sequences and reference sequences from a pre-existing database to determine which taxon a sequence may belong to (Sedlar *et al.*, 2017). Publicly available programs including RDP Classifier and NCBI BLAST are commonly used to identify closely related organisms to sequences of interest, allowing researchers to directly assess the taxonomic diversity of a particular sample. Taxonomic dependent binning of metagenomic sequences can also be used for the facilitation of genome assembly and the evaluation of gene homology that may exist among different taxonomic groups (Tringe *et al.*, 2005; Frank *et al.*, 2016). Functionality can also be inferred to binned reads that are closely related to a characterized organism. Limitations in this approach include the extended length of time it can take to compare large sequence datasets against reference databases, especially when using alignment

methods, as well as the incomplete reference database used to compare sample sequences to sequences of known taxonomy (Sedlar *et al.*, 2017). Further, limitations in reference databases are particularly challenging when sampling and evaluating environments where few microorganisms have been characterized.

Taxonomic independent strategies work by applying clustering techniques based on DNA sequence features without relying on reference sequences to group microorganisms into different categories. This approach is commonly used to study composition and diversity in complex microbiomes, such as the rumen, where the majority of the microorganisms are uncharacterized (Kim *et al.*, 2011b). Taxonomic independent binning of sequenced single-gene markers, however, gives little information regarding the functionality of the microorganisms being sampled. As a result, further investigation is needed in order to determine the metabolic potential of microbes binned independent of taxonomy.

Cellulose: Structure and Microbial Breakdown

Cellulose, the most abundant biopolymer found on earth, consists of repeated D-glucose subunits linked by β (1-4) glycosidic bonds. A single molecule of cellulose can be comprised of hundreds or up to thousands of monomeric glucose subunits, making it one of the longest organic polymers found in nature (Lee *et al.*, 2014). In the primary cell walls of plants, cellulose can be assembled into long, rigid microfibrils that allow the plant to control cell expansion, and serves as a key component in growth and development (Thomas *et al.*, 2013). While it is the major structural component of plant cell walls, cellulose can also be found in certain species of algae, fungi, as well as select bacterial biofilms (Serra *et al.*, 2013).

Enzymes responsible for cellulose breakdown have yet to be identified in any animal genome, but they are expressed by bacteria, protozoa, and fungi. In the rumen, a number of different cellulases involved in the hydrolysis of cellulose have been characterized, and most ruminal cellulases belong to a broad family of enzymes known as glycoside hydrolases (GH). Within the 135 groups of GHs (www.cazy.org), members of this family can be divided into three broad functional groups based on where within the cellulose polymer the glycosidic bond is cleaved by the enzyme: endocellulases, exocellulases, and β-glucosidases. In bacteria, these three types are expressed as extracellular enzymes that work synergistically in the conversion of long polysaccharide chains to D-glucose subunits (Krause *et al.*, 2003).

As feed is ingested by the host ruminant, polymers such as cellulose become available to ruminal microorganisms. Endocellulases function as the initial biochemical tools in cellulose breakdown, by scission of β (1-4) glycosidic bonds at random positions within the cellulose polymer, yielding cellu-oligosaccharides possessing new chain ends. The non-reducing ends of the newly formed cellu-oligosaccharides then become available for the active site in exocellulases (Krause *et al.*, 2003). Typically, exocellulases can cleave the β (1-4) glycosidic bond 2-4 subunits from the non-reducing end, forming cellobiose as their primary product (Zverlov *et al.*, 2005). Cellobiose is a disaccharide consisting of two D-glucose monomers linked by a β (1-4) glycosidic bond that can subsequently be cleaved by β-glucosidases to release the monomeric glucose subunits (Krause *et al.*, 2003). Soluble glucose can then be absorbed by the host or it can be transported across bacterial membranes for intracellular fermentation into VFAs (Thauer *et al.*, 2008).

Most known fiber-degrading bacteria possess the ability to adhere directly to the plant cell wall, allowing them better access to the soluble products from cellulose hydrolysis (Schwarz, 2001). Adhesion requires the expression of other proteins, such as dockerins and cohesion domains, which facilitate attachment of the enzymatic complex to the bacterial surface, allowing the bacterial membrane to be in close proximity of the substrate being metabolized. Bacterial adhesion to the plant cell wall can be impeded by factors including: competition with other fiber-adhering microorganisms, surface area of the substrate available, pH, temperature, and glycocalyx condition (Miron *et al.*, 2001). Bacterial species shown to adhere directly to plant fibers include *Fibrobacter succinogenes*, *Ruminococcus albus*, and *Ruminococcus flavefaciens* (Latham *et al.*, 1978; Miron & Ben-Ghedalia, 1993; Bayer *et al.*, 1998). Some microorganisms have evolved specialized extracellular enzymatic complexes, known as cellulosomes, which allow the cell to both physically bind with the plant fiber and catalyze the breakdown of cellulose. These multi-subunit complexes allow bacteria to attach to cellulose via the cellulose binding domain, permitting the GHs of the catalytic domain to hydrolyze the available β (1-4) glycosidic bonds (Steenbakkers *et al.*, 2001).

In addition to cellulosomes, bacteria as a group have evolved other mechanisms that allow them to break down complex carbohydrates. In the human gastrointestinal tract, species belonging to the phylum *Bacteroidetes* have been shown to possess operons, referred to as polysaccharide utilization loci (PULs), that express components for the sensing, binding, and cleavage of polysaccharides, as well as the import of oligosaccharides into the periplasm, and their depolymerization into monosaccharides (Terrapon & Henrissat, 2014). El Kaoutari *et al*. (2013) have described how certain

Bacteroidetes species have evolved a diversified repertoire of PULs, with certain species possessing dozens of different PULs within their respective genome. Due to the high abundance of *Bacteroidetes* in the gastrointestinal tract of mammals, other metagenomic studies have focused on identifying PULs expressed by microorganisms in other host species, such as ruminant animals, leading to further insight on cellulolytic mechanisms used by microorganisms in the rumen (Pope *et al.*, 2010; Hess *et al.*, 2011; Pope *et al.*, 2012; Rosewarne *et al.*, 2014).

Other non-catalytic proteins have been identified as tools that bacteria use to aid in the metabolism of cellulose microfibrils. Originally detected in plants, expansins are needed to loosen tightly woven cellulose of plant cell walls to allow for growth and development (Cosgrove, 2000; Arantes & Saddler, 2010). The mechanism for these non-enzymatic proteins has been proposed as a disrupting action of the non-covalent bond between the matrix polysaccharide and the associated cellulose microfibrils, allowing the polysaccharides to be more accessible to cellulolytic enzymes (McQueen-Mason & Cosgrove, 1994; Cosgrove, 2000). While expansins found in plants and expansin-like proteins recently found in bacteria and fungi share functional similarities, the direct relationship between these proteins is unclear. The non-plant homologues have been grouped together in the expansin-like family X (Kende *et al.*, 2004).

In the cell walls of plants, cellulose is rarely a standalone polymer. Other structural compounds such as aromatic esters, pectin, hemicellulose, and lignin are typically associated with plant cell walls, rendering access difficult for the active site of many cellulase enzymes (Krause *et al.*, 2003). This decreased accessibility to cellulose by cellulase enzymes has been shown to affect fiber digestibility and was demonstrated by

Zadrazil *et al*. (1995), where wheat straw was incubated with *Pleurotis* spp., a fungal species known to express ligninolytic enzymes, and found that *in vitro* digestibility (dry matter disappearance) improved with the addition of the fungi.

Research Objective

Bacteria have been found to be the most diverse and abundant cell type in rumen environments. Based on available sequence metadata generated from culture-independent methods, it has been estimated that there are at least 7,400 ruminal bacterial species. While several enzymes involved in cellulose breakdown as well as some of the denizens of the rumen performing this elaborate task have been identified, it has been hypothesized that only 5% of ruminal bacteria are available as cultured strains (Kim *et al.*, 2011a). The advent of culture-independent methods has unveiled a more comprehensive approach for ruminal community analysis that allows us to identify a large number of microorganisms while eliminating biases from culture-dependent techniques. In this context, the objective of the research presented in this thesis was to identify previously uncharacterized ruminal bacteria based on their ability to grow using cellulose as a selective substrate. Identification of these bacteria could allow for a greater understanding of ruminal cellulose metabolism, with the potential to offer new insight on strategies to enhance feed efficiency in ruminant livestock.

Rationale and Approach

Combining selection-based culturing with the holistic approach of culture-independent methods represents a promising strategy to identify previously uncharacterized ruminal bacteria involved in cellulose metabolism. By adding cellulose

as the only substrate to ruminal fluid cultures, microorganisms able to degrade this polysaccharide for the production of energy, or are able to benefit from by-products of cellulolytic metabolism, will thrive and grow in abundance, whereas those without these abilities should remain stable or decrease in population. Since the majority of ruminal bacteria have yet to be characterized as species, OTU clustering, a taxonomic-independent binning approach, was used to group microorganisms at the species level. The use of OTU clustering allows for monitoring shifts in the abundance of uncharacterized bacterial groups in rumen cultures as a result of selection with cellulose. While we cannot conclude that an OTU that increases in abundance does in fact possess cellulolytic capabilities, results from this study serve as a key initial step towards the characterization of currently unknown bacteria that perform cellulose breakdown in the rumen.

Literature Cited

Achenbach LA, Carey J & Madigan MT (2001) Photosynthetic and phylogenetic primers for detection of anoxygenic phototrophs in natural environments. *Applied and Environmental Microbiology* **67**: 2922-2926.

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW & Nielsen PH (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* **31**: 533-538.

Angelidaki I, Karakashev D, Batstone DJ, Plugge CM & Stams AJ (2011) Biomethanation and its potential. *Methods in Enzymology* **494**: 327-351.

Arantes V & Saddler JN (2010) Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnology for Biofuels* **3**: 4.

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

Bayer EA, Chanzy H, Lamed R & Shoham Y (1998) Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology* **8**: 548-557.

Bodirsky BL, Rolinski S, Biewald A, Weindl I, Popp A & Lotze-Campen H (2015) Global food demand scenarios for the 21(st) century. *PloS One* **10**(11): e0139201.

Bruns KW, Pritchard RH & Boggs DL (2005) The effect of stage of growth and implant exposure on performance and carcass composition in steers. *Journal of Animal Science* **83**: 108-116.

Byerly TC (1967) Efficiency of feed conversion. *Science* **157**: 890-895.

Case RJ, Boucher Y, Dahllof I, Holmstrom C, Doolittle WF & Kjelleberg S (2007) Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Applied and Environmental Microbiology* **73**: 278-288.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR & Tiedje JM (2014) Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research* **42**: D633–D642.

Cosgrove DJ (2000) Loosening of plant cell walls by expansins. *Nature* **407**: 321-326.

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

Dawson IGJ & Johnson JEV (2014) Growing Pains: How Risk Perception and Risk Communication Research Can Help to Manage the Challenges of Global Population Growth. *Risk Analysis* **34**: 1378-1390.

Denman SE & McSweeney CS (2015. 3:9.1–9.19) The early impact of genomics and metagenomics on ruminal microbiology. *Annual Review of Animal Biosciences* **3**: 9.1-9.19.

Droge J & McHardy AC (2012) Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Briefings in Bioinformatics* **13**: 646-655.

Edwards EJ, McEwan RN, Travis JA & Wallace RJ (2004) 16S rDNA library based analysis of ruminal bacterial diversity. *Antonie Van Leeuwenhoek* **86**: 263-281.

El Kaoutari A, Armougom F, Gordon JI, Raoult D & Henrissat B (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Reviews Microbiology* **11**: 497-504.

Epstein SS (2013) The phenomenon of microbial uncultivability. *Current Opinion in Microbiology* **16**: 636-642.

Ferrer M, Golyshina OV, Chernikova TN*, et al.* (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology* **7**: 1996-2010.

Ferry JG (2010) The chemical biology of methanogenesis. *Planet Space Science* **58**: 1775-1783.

Fouts DE, Szpakowski S, Purushe J, Torralba M, Waterman RC, MacNeil MD, Alexander LJ & Nelson KE (2012) Next generation sequencing to define prokaryotic and fungal diversity in the bovine rumen. *PloS One* **7**: e48289.

Frank JA, Pan Y, Tooming-Klunderud A, Eijsink VG, McHardy AC, Nederbragt AJ & Pope PB (2016) Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. *Scientific Reports* **6**: 25373.

Handelsman J, Rondon MR, Brady SF, Clardy J & Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology* **5**: R245-249.

Henderson G, Cox F, Ganesh S, Jonker A, Young W & Janssen PH (2015) Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports* **5**: 14567.

Hendrickx B, Dejonghe W, Faber F, Boenne W, Bastiaens L, Verstraete W, Top EM & Springael D (2006) PCR-DGGE method to assess the diversity of BTEX mono-oxygenase genes at contaminated sites. *FEMS Microbiological Ecology* **55**: 262-273.

Hess M, Sczyrba A, Egan R*, et al.* (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science (New York, NY)* **331**: 463–467.

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

Johnson DR, Goldschmidt F, Lilja EE & Ackermann M (2012) Metabolic specialization and the assembly of microbial communities. *ISME Journal* **6**: 1985-1991.

Jose VL, Appoothy T, More RP & Arun AS (2017) Metagenomic insights into the rumen microbial fibrolytic enzymes in Indian crossbred cattle fed finger millet straw. *AMB Express* **7**: 13.

Kende H, Bradford K, Brummell D*, et al.* (2004) Nomenclature for members of the expansin superfamily of genes and proteins. *Plant Molecular Biology* **55**: 311-314.

Kerr BJ, Kellner TA & Shurson GC (2015) Characteristics of lipids and their feeding value in swine diets. *Journal of Animal Science and Biotechnology* **6**: 30.

Kim M, Morrison M & Yu Z (2011a) Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiological Ecology* **76**: 49–63.

Kim M, Morrison M & Yu Z (2011b) Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of Microbiological Methods* **84**: 81-87.

Krause DO & Russell JB (1996) How many ruminal bacteria are there? *Journal of Dairy Science* **79**: 1467-1475.

Krause DO, Denman SE, Mackie RI, Morrison M, Rae AL, Attwood GT & McSweeney CS (2003) Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. *FEMS Microbiology Reviews* **27**: 663-693.

Lane D, Pace B, Olsen GJ, Stahl DA, Sogin ML & Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the USA* **82**: 6955-6959.

Latham MJ, Brooker BE, Pettipher GL & Harris PJ (1978) Ruminococcus flavefaciens Cell Coat and Adhesion to Cotton Cellulose and to Cell Walls in Leaves of Perennial Ryegrass (Lolium perenne). *Applied and Environmental Microbiology* **35**: 156-165.

Lee HV, Hamid SB & Zain SK (2014) Conversion of lignocellulosic biomass to nanocellulose: structure and chemical process. *The Scientific World Journal* **2014**: 631013.

Li RW, Connor EE, Li C, Baldwin RL & Sparks ME (2011) Characterization of the rumen microbiota of preruminant calves using metagenomic tools. *Environmental Microbiology* **14**: 129-139.

Margesin R & Miteva V (2011) Diversity and ecology of psychrophilic microorganisms. *Research in Microbiology* **162**: 346-361.

McQueen-Mason S & Cosgrove DJ (1994) Disruption of hydrogen bonding between plant cell wall polymers by proteins that induce wall extension. *Proceedings of the National Academy of Sciences of the USA* **91**: 6574-6578.

McSweeney C & Mackie R (2012) Micro-organisms and ruminant digestion: state of knowledge, trends and future prospects. Background study paper No. 61. Food Agrie. Org. United Nations, Rome, Italy.

Miron J & Ben-Ghedalia D (1993) Digestion of cell-wall monosaccharides of ryegrass and alfalfa hays by the ruminal bacteria Fibrobacter succinogenes and Butyrivibrio fibrisolvens. *Canadian Journal of Microbiology* **39**: 780-786.

Miron J, Ben-Ghedalia D & Morrison M (2001) Invited review: adhesion mechanisms of rumen cellulolytic bacteria. *Journal of Dairy Science* **84**: 1294-1309.

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

Muya MC, Nherera FV, Miller KA, Aperce CC, Moshidi PM & Erasmus LJ (2015) Effect of Megasphaera elsdenii NCIMB 41125 dosing on rumen development, volatile fatty acid production and blood β-hydroxybutyrate in neonatal dairy calves. *Journal of Animal Physiology and Animal Nutrition* **99**: 913-918.

Naas AE, Mackenzie AK, Mravec J, Schückel J, Willats WG, Eijsink VG & Pope PB (2014) Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio* **5**: e01401.

Nicolaisen MH & Ramsing NB (2002) Denaturing gradient gel electrophoresis (DGGE) approaches to study the diversity of ammonia-oxidizing bacteria. *Journal of Microbiological Methods* **50**: 189-203.

Patel JB (2001) 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Molecular Diagnosis* **6**: 313-321.

Patil KR, Haider P, Pope PB, Turnbaugh PJ, Morrison M, Scheffer T & McHardy AC (2011) Taxonomic metagenome sequence assignment with structured output models. *Nature Methods* **8**: 191-192.

Pham VH & Kim J (2012) Cultivation of unculturable soil bacteria. *Trends in Biotechnology* **30**: 475-484.

Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC, Morrison M & Eijsink VG (2012) Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PloS One* **7**: e38571.

Pope PB, Denman SE, Jones M*, et al.* (2010) Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other

herbivores. *Proceedings of the National Academy of Sciences of the USA* **107**: 14793-14798.

Puspita ID, Kamagata Y, Tanaka M, Asano K & Nakatsu CH (2012) Are uncultivated bacteria really uncultivable? *Microbes and Environments* **27**: 356-366.

Raes J & Bork P (2008) Molecular eco-systems biology: towards an understanding of community function. *Nature Reviews Microbiology* **6**: 693-699.

Reddy C & Peck H (1978) Electron transport phosphorylation coupled to fumarate reduction by H2 and Mg2+ dependent adenosine triphosphate activity in extracts of the rumen anaerobe *Vibrio succinogenes*. *Journal of Bacteriology* **134**: 982-991.

Rey M, Enjalbert F & Monteils V (2012) Establishment of ruminal enzyme activities and fermentation capacity in dairy calves from birth through weaning. *Journal of Dairy Science* **95**: 1500-1512.

Rey M, Enjalbert F, Combes S, Cauquil L, Bouchez O & Monteils V (2014) Establishment of ruminal bacterial community in dairy calves from birth to weaning is sequential. *Journal of Applied Microbiology* **116**: 245-257.

Rosewarne CP, Pope PB, Cheung JL & Morrison M (2014) Analysis of the bovine rumen microbiome reveals a diversity of Sus-like polysaccharide utilization loci from the bacterial phylum Bacteroidetes. *Journal of Industrial Microbiology and Biotechnology* **41**: 601-606.

Rosling A, Cox F, Cruz-Martinez K, Ihrmark K, Grelet GA, Lindahl BD, Menkis A & James TY (2011) Archaeorhizomycetes: unearthing an ancient class of ubiquitous soil fungi. *Science* **333**: 876-879.

Schwarz WH (2001) The cellulosome and cellulose degradation by anaerobic bacteria. *Applied Microbiology and Biotechnology* **56**: 634-649.

Sedlar K, Kupkova K & Provaznik I (2017) Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics. *Computational and Structural Biotechnology Journal* **15**: 48-55.

Serra DO, Richter AM & Hengge R (2013) Cellulose as an architectural element in spatially structured Escherichia coli biofilms. *Journal of Bacteriology* **195**: 5540-5554.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM & Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the USA* **103**: 12115-12120.

Steenbakkers PJM, Li XL, Ximenes EA, Arts JG, Chen H, Ljungdahl LG & Op den Camp HJM (2001) Noncatalytic docking domains of cellulosomes of anaerobic fungi. *Journal of Bacteriology* **183**: 5325-5333.

Terrapon N & Henrissat B (2014) How do gut microbes break down dietary fiber? *Trends in Biochemical Sciences* **39**: 156-158.

Thauer RK, Kaster A-K, Seedorf H, Buckel W & Hedderich R (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature reviews Microbiology* **6**: 579-591.

Thomas LH, Forsyth VT, Sturcova A, Kennedy CJ, May RP, Altaner CM, Apperley DC, Wess TJ & Jarvis MC (2013) Structure of cellulose microfibrils in primary cell walls from collenchyma. *Plant Physiology* **161**: 465-476.

Thornton PK (2010) Livestock production: recent trends, future prospects. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **365**: 2853-2867.

Throback IN, Enwall K, Jarvis A & Hallin S (2004) Reassessing PCR primers targeting nirS, nirK and nosZ genes for community surveys of denitrifying bacteria with DGGE. *FEMS Microbiology Ecology* **49**: 401-417.

Tringe SG, von Mering C, Kobayashi A*, et al.* (2005) Comparative metagenomics of microbial communities. *Science* **308**: 554-557.

Valin H, Sands RD, van der Mensbrugghe D*, et al.* (2014) The future of food demand: understanding differences in global economic models. *Agricultural Economics* **45**: 51-67.

Van Soest PJ (1994) Nutrional ecology of the ruminant (second edition). 476 pages.

Varel VH, Yen JT & Kreikemeier KK (1995) Addition of cellulolytic clostridia to the bovine rumen and pig intestinal tract. *Applied and Environmental Microbiology* **61**: 1116-1119.

Vinje H, Almoy T, Liland KH & Snipen L (2014) A systematic search for discriminating sites in the 16S ribosomal RNA gene. *Microbial Informatics and Experimentation* **4**: 2.

von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, Ward N & Bork P (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**: 1126-1130.

Wang L, Hatem A, Catalyurek UV, Morrison M & Yu Z (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PloS One* **8**: e78507.

Wang W, Li C, Li F*, et al.* (2016) Effects of early feeding on the host rumen transcriptome and bacterial diversity in lambs. *Scientific Reports* **6**: 32479.

Weimer PJ, Russell JB & Muck RE (2009) Lessons from the cow: what the ruminant animal can teach us about consolidated bioprocessing of cellulosic biomass. *Bioresource Technology* **100**: 5323-5331.

Weimer PJ, Stevenson DM & Mertens DR (2010) Shifts in bacterial community composition in the rumen of lactating dairy cows under milk fat-depressing conditions. *Journal of Dairy Science* **93**: 265-278.

Weimer PJ, Stevenson DM, Mantovani HC & Man SLC (2010) Host specificity of the ruminal bacterial community of the dairy cow following near-total exchange of ruminal contents. *Journal of Dairy Science* **93**: 5902-5912.

Yeoman CJ & White BA (2014) Gastrointestinal tract microbiota and probiotics in production animals. *Annual Review of Animal Biosciences* **2**: 469-486.

Yue ZB, Li WW & Yu HQ (2013) Application of rumen microorganisms for anaerobic bioconversion of lignocellulosic biomass. *Bioresource Technology* **128**: 738-744.

Zadrazil F, Puniya AK & Singh K (1995) Pilot scale reactor for biological treatment of lignocellulosics for animal feed production. . *Indian Journal of Dairy Science* **48**: 110-117.

Ze X, Le Mougen F, Duncan SH, Louis P & Flint HJ (2013) Some are more equal than others: the role of "keystone" species in the degradation of recalcitrant substrates. *Gut Microbes* **4**: 236-240.

Zhou F, Olman V & Xu Y (2008) Barcodes for genomes and applications. *BMC Bioinformatics* **9**: 546.

Zverlov VV, Schantz N & Schwarz WH (2005) A major new component in the cellulosome of Clostridium thermocellum is a processive endo-beta-1,4-glucanase producing cellotetraose. *FEMS Microbiology Letters* **249**: 353-358.

Chapter 2

IDENTIFICATION OF CANDIDATE CELLULOSE UTILIZING BACTERIA FROM THE RUMEN OF BEEF CATTLE USING A BATCH CULTURE APPROACH

Abstract

The microbial-driven process of converting cellulosic biomass into utilizable energy is a defining feature of ruminant animals. Due to the complexity of rumen microbial communities, the vast majority of rumen microorganisms remain uncharacterized, which include bacteria capable of metabolizing cellulose. To gain further insight, selection-based batch culturing from bovine rumen fluid was used to identify previously uncharacterized cellulose-utilizing rumen bacteria. A 16S rRNA-based population analysis was used to identify rumen bacteria that were enriched within 14 days of culturing. A total of 116,111 to 170,626 high quality, non-chimeric sequence reads per sample were used for operational taxonomic unit (OTU) clustering. Using this approach, five different candidate cellulose-utilizing species-level OTUs were identified. Four of the enriched OTUs showed increased levels ranging between 69 and 445-fold increases compared to their respective rumen inocula, which represented 15.9% to 41.3% of reads in samples supplemented with cellulose, while their relative abundance in control samples ranged from 0.08% to 0.25%. One OTU corresponded to a known species (*Ruminococcus flavefaciens*), while another OTU was predicted to be an uncultured species of a known genus (*Ethanoligenens* sp.), and two were assigned to the family *Ruminococcaceae*. The remaining OTU (*Prevotella* sp.) was found in high abundance in the ruminal inoculum (33.6%) and was maintained at relatively high levels in the presence of cellulose (15.6%). While future research is necessary to further characterize their involvement in cellulose metabolism, identification of these bacteria will provide additional insights towards a better understanding of ruminal cellulose metabolism.

Introduction

As the main component of plant cell walls and sink for photosynthesis products, cellulose represents one of the most abundant organic polymers and reserve monosaccharides on earth (Naas *et al.*, 2014). Consequently, herbivory has successfully evolved in certain animal groups to take advantage of this abundant source of energy. However, since genes encoding cellulose-metabolizing enzymes have yet to be identified in the genomes of mammals or other animals, herbivores typically rely on gastrointestinal microbial symbionts to digest plant-cell wall polysaccharides (Van Soest, 1994; McSweeney & Mackie, 2012; Yue *et al.*, 2013).

Since cellulose and other components of plant fibers represent recalcitrant substrates even for microorganisms that can metabolize them, two main strategies for mammalian herbivory have successfully evolved to maximize their digestion efficiency. In hindgut fermenters, the distal segments of the gastrointestinal tract are the most developed, resulting in increased intestinal volume to favor retention time and maintenance of high microbial numbers (Macfarlane & Macfarlane, 1993). In contrast, ruminants have developed a compartmentalized stomach, of which the largest segment, the rumen, hosts the microbial symbionts responsible for digesting the feed consumed by the animal (Van Soest, 1994). Ruminants represent a highly successful biological group that has thrived across a wide range of habitats, and also played a central role throughout human history. Even to this day, the ability of domesticated ruminants to transform inedible plant biomass into protein-rich products that can be consumed by humans is expected to greatly contribute in meeting the demands of a rapidly growing and urbanizing global population (Thornton, 2010).

Rumen microbial symbionts form complex ecosystems that include a diverse array of members from the three domains of life: bacteria, archaea, protozoa, and fungi. Due to their anaerobic lifestyle, rumen microorganisms tend to have specialized metabolic functions, and to form complex trophic relationships with other community members to fulfill their metabolic needs (Johnson *et al.*, 2012). Since bacteria are found at the highest cell density and represent the most diverse group in terms of genetic and metabolic functions, they have been the subjects of most investigations on rumen microbial function (McSweeney & Mackie, 2012). While a number of isolates from major bacterial phyla such as *Bacteroidetes*, *Firmicutes*, and *Fibrobacteres* which include well characterized representatives such as *Prevotella ruminicola*, *Ruminococcus albus*, and *Fibrobacter succinogenes*, respectively, have been isolated, they may as a group represent only 6% of rumen cellulose utilizing bacterial species (Russell *et al.*, 2009; Creevey *et al.*, 2014). In support of this estimate, a number of rumen metagenomics studies have reported an abundance of previously unidentified carbohydrate-active genes, suggesting a wealth of uncharacterized bacteria involved in cellulose catabolism that have yet to be isolated (Brulc *et al.*, 2009; Hess *et al.*, 2011; Wang *et al.*, 2013). To gain further insight into uncharacterized rumen bacteria that are involved in cellulose breakdown, we present in this report five candidate cellulose metabolizing bacteria that were identified by enrichment from rumen fluid using a batch culture approach.

Materials and Methods

Rumen fluid collection and culturing

All animal procedures were approved by the Institutional Animal Care and Use Committee at South Dakota State University. Samples were collected between December

2015 and October 2016 from fistulated Simental × Angus cows maintained at the South Dakota State University Cow and Calf Education and Research Facility. Cattle were fed fiber-based diets (i.e. pasture hay and haylage). For each culture experiment, 12 L of fresh rumen fluid were collected by manually squeezing the liquid fraction from digesta, and the samples were stored in insulated containers during transit. For culture setup, which took place within one hour of sample collection, rumen fluid stored in separate containers were combined, mixed, and distributed amongst five bioreactors (Chemglass) with volumes of approximately 2.2 L and 0.8 L for each culture and biogas (headspace), respectively. Cellulose (20g/L, Sigma cat# C8002) was added to three replicate cultures, while the remaining two bioreactors were not supplemented with any substrate. Cultures were maintained at a temperature of 38°C, with stirring at 150 rpm. Each bioreactor was equipped with a flexible plastic tube to allow for the release of excess biogas and prevent pressure build-up. The pH of cultures was measured at designated time points during the experimental period using a Portable Meter Kit (Oakton Instruments, cat# 35613-24), which was calibrated prior to each measurement.

Microbial DNA extraction and PCR amplification of the 16S rRNA gene

Microbial DNA was isolated from rumen and culture samples using the repeated bead beating and column method, as described by Yu and Morrison (2004). The V1-V3 region of bacterial 16S rRNA gene sequences was PCR-amplified using the 27F forward (Edwards *et al.*, 1989) and 519R reverse primer pair (Lane *et al.*, 1985). PCR reactions were performed with the Phusion *Taq* DNA polymerase (Thermo Scientific) under the following conditions: hot start (4 min, 98 ^{o}C), followed by 35 cycles of denaturation (10s, 98 ^{o}C), annealing (30s, 50 ^{o}C) and extension (30 s, 72 ^{o}C), then ending with a final

extension period (10 min, 72 ºC). PCR products were separated by agarose gel electrophoresis, and amplicons of the expected size (~500bp) were excised for gel purification using the QiaexII Gel extraction kit (Qiagen). For each sample, at least 400 ng of amplified DNA were submitted to Molecular Research DNA Lab (MRDNA) (Shallowater, TX, USA) for sequencing with the Illumina MiSeq 2X300 platform to generate overlapping paired end reads.

Computational analysis of PCR-generated 16S rRNA amplicon sequences

Raw bacterial 16S rRNA gene V1-V3 amplicon sequences were assembled as contigs from overlapping MiSeq 2x300 paired-end reads from the same flow cell clusters. Reads were then selected using custom Perl scripts to meet the following criteria: presence of both 27F (forward) and 519R (reverse) primer nucleotide sequences, length between 400 and 580 nt, and a minimal quality threshold of no more than five nucleotides with a Phred quality score lower than 15. Using custom Perl scripts, these selected reads were then aligned, and clustered into operational taxonomic units (OTUs) at a species-level genetic distance cutoff of 5% sequence dissimilarity (St-Pierre & Wright, 2015).

OTUs were screened for sequence artifacts using three separate methods. Chimeric sequences were first identified with the chimera.uchime command from the MOTHUR open source software package (Schloss *et al.*, 2009). Secondly, OTUs with fewer than 500 reads that were not assigned to a known bacterial phylum by RDP Classifier ("unclassified bacteria" at an 80% bootstrap cutoff) (Wang *et al.*, 2007) were designated as artifacts. Finally, the integrity of the 5′ and 3′ ends of OTUs was tested using a database alignment search based approach; when compared to their closest match of equal or longer sequence length from the NCBI nt database found by BLAST

(Altschul *et al.*, 1997), OTUs with more than five nucleotides missing from the 5′ or 3′ end of their respective alignments were designated as artifacts. After removal of chimeras and artifacts, taxonomic assignment of valid OTUs was performed using a combination of RDP Classifier and BLAST. The List of Prokaryotic Names with Standing in Nomenclature (LPSN) was also consulted for information on valid species belonging to various taxa of interest.

Volatile fatty acid (VFA) analysis

Rumen or culture samples collected for VFA composition analysis were mixed with metaphosphoric acid (25%) at a 4:1 ratio, then stored frozen (-20°C) until analysis was performed. Prior to VFA analysis, sampled were thawed, then centrifuged at 16,000 x g (1 min) to remove particulates. VFA concentrations were determined using a Trace 1310 Gas Chromatograph (Thermo Scientific).

Results

Bacterial community structure in uncultured rumen samples

Prior to selection with cellulose, the composition of the bacterial populations from each rumen inoculum of interest was analyzed. Bacterial diversity ranged from 5,480 to 6,756 species-level OTUs (Figure 2.1). Taxonomic analysis revealed that the phylum *Bacteroidetes* was the most abundant in all three samples (51.6%-62.3% of reads per sample), with *Prevotellaceae* identified as the most highly represented family (20.2%-38.2% of reads per sample). *Firmicutes* was the second most abundant phylum, ranging from 25.6% in sample 1 to 42.1% in sample 3, with *Clostridiales* showing the highest representation within this group. *Lachnospiraceae* was the most predominant *Firmicutes*

family in samples 1 and 2, with abundances of 10.3% and 11.8%, respectively, while unassigned or unclassified was the most abundant group of *Clostridiales* in sample 3 (15.7%).

Rumen samples were further characterized through identification of well represented OTUs, defined as representing at least 1% of sequences in one or more samples. In sample 1, seven OTUs were found to meet this criteria. Its OTU with the highest abundance was assigned to the phylum *Fibrobacteres* (4.7%), while the second most abundant (3.9%) did not correspond to any currently defined phylum (Table 2.3). The remaining OTUs were represented by the phyla *Bacteroidetes* (three OTUs) and *Planctomycetes* (two OTUs). No OTUs from sample 2 were found to have an abundance greater than 1%. In sample 3, OTU (LOE10-04095), assigned to *Prevotella*, was the most highly represented in our study (33.6%), while the second OTU, with an abundance of 1.0%, belonged to the candidate phylum SR1 (Table 2.5).

Rumen bacterial community composition after 7 days in culture with cellulose

To identify novel rumen bacteria that can metabolize cellulose, changes in taxonomic distribution and OTU abundances were monitored in rumen cultures supplemented with cellulose, in comparison to their sample of origin. An overall decrease in *Bacteroidetes* and increase in *Firmicutes* was observed in all cultures after 7 days in the presence of cellulose (Table 2.2). Indeed, in samples 1 and 2, *Prevotellaceae* showed a decrease in abundance of 16.5%-18.0%, whereas unclassified *Bacteroidales* exhibited a reduction of 10.4% in sample 3, which were the sharpest declines observed in their respective samples. *Ruminococcaceae* (*Firmicutes*) showed an increase of 9.7% and 10.0% for samples 1 and 3, respectively, in contrast to sample 2, whose representation of

unclassified *Clostridiales* was higher by 7.3% in sample 2. While families of the phylum *Firmicutes* overall appeared to increase in the presence of cellulose, a decrease in *Lachnospiraceae* of 2.2% and 3.7% was observed in samples 1 and 2, respectively.

After seven days in culture, certain OTUs were found in higher abundance in response to cellulose. For sample 1, the greatest increase in relative abundance was seen in two OTUs, which were assigned to *Saccharibacteria_genera_incertae_sedis* of the candidate phylum TM7 and unclassified *Ruminococcaceae*, respectively, changing from 0.4% to 7.0%, and from 0.3% to 4.6% (Table 2.3). Ten other OTUs showed relative abundances greater than 1% in sample 1, with the majority belonging to *Firmicutes* (four OTUs) and *Bacteroidetes* (three OTUs), while *Proteobacteria*, *Planctomycetes*, and unclassified bacteria were each represented by one OTU. Three OTUs (C1-00068, C1-16337, and C1-05580) well represented in the original rumen samples displayed only minor fluctuations in abundance after seven days in culture.

In sample 2, only two well represented OTUs were found on d7 (Table 2.4). One OTU, belonging to the genus *Fibrobacter*, showed the largest increase in response to cellulose (0.7% to 7.7%), whereas the other, assigned to the genus *Prevotella*, increased from 0.1% to 1.6%. Five well represented OTUs were identified from sample 3 after seven days in culture (Table 2.5). Two OTUs (LOE4-02289 and LOE2-14324), corresponding to uncharacterized species of *Prevotella* (0.1% to 4.2%) and *Alphaproteobacteria* (0.2% to 2.5%), respectively, showed the highest increase. While its abundance decreased after seven days in culture with cellulose, OTU (LOE10-04095) remained well represented in two replicates (1.4% and 10.9%). Two OTUs (LOE2-14329 and LOE4-22899) assigned to *Firmicutes* were also found to be well represented.

Identification of candidate cellulose utilizing rumen bacteria from enriched cultures

After 14 days in cultures supplemented with cellulose, five OTUs were found to be in greater abundance across the three samples (Figure 2.2). OTU (C10-00008) increased from 0.18% in rumen sample 1 to 51% and 72% in two cellulose-supplemented cultures, an increase of 240-340 fold. In comparison, control d14 replicates showed a mean relative abundance of 0.26% for this OTU. An uncultured bacterial species identified in feces of gorilla and flying fox (EU775520.1) was found to be almost identical (99%) to OTU C10-00008 (Ley *et al.*, 2008). *Ruminococcus flavefaciens* was identified as its closest valid relative (GU999991.1), with a sequence identity of 95%.

Culturing of sample 2 in the presence of cellulose resulted in the enrichment of three OTUs, all assigned to *Ruminococcaceae*. The abundance of OTU LOD11-02512 increased from 0.09% to 35.2%, and 9.1% in two cellulose-supplemented cultures after 14 days, respectively, while the mean relative abundance for the control replicates at the same time point was 0.1%. This OTU was assigned to an unidentified genus within *Ruminococcaceae*. Its closest match in sequence (97%) was an uncultured *Clostridium* bacterium isolated from a solid state anaerobic digester containing maize silage (JX099828.1) (Sträuber *et al.*, 2012), while its closest valid relative (93%) was found to be *Caproiciproducens galactitolivorans* (NR145929.1). The second main OTU from sample 2, LOD10-23342, also an unclassified *Ruminococcaceae* genus, increased to 26.5% and 20.5% in two of the replicate cellulose supplemented cultures, compared to 0.1% in the d14 control cultures. This OTU was most closely related (95%) to an uncultured bacterium found in springs and wells of the Mojave Desert (KF836228.1), with *Clostridium cellulosi* (FJ465164.1) as its closest valid relative (93%). The third

OTU from sample 2, assigned to the genus *Ethanoligenens* by RDP Classifier, was enriched in all experimental cultures (6.0%, 22.1%, and 24.0%, respectively). A BLAST search revealed that its highest identity (99%) was to an uncultured bacterium (GQ453659.1) isolated from decomposed municipal solid waste, and that *Clostridium cellulosi* (FJ465164.1) was its closest relative (94%).

Enrichment in response to cellulose from sample 3 resulted in the identification of OTU LOE10-04095, which was assigned to the genus *Prevotella*. Found in high abundance in its original rumen sample (33.6%), this OTU was maintained in two of the experimental replicates on D14 (10.6% and 35.9%). In contrast, its mean relative abundance was only 0.08% in the D14 unsupplemented control. OTU LOE10-04095 was found to be very close (99%) to an uncultured rumen bacterium identified in cattle fed a diet containing cashew nut shells (AB616459.1), while it was only 91% identical in sequence to its closest valid species (*Prevotella melaninogenica*, LT684184.1).

Determination of VFA levels in rumen and cultures

As a means to determine the potential fermentation end product of the enriched OTUs, the concentrations of acetate, propionate, and butyrate were determined (Table 2.6). In sample 1, concentrations of all three VFAs analyzed increased beyond the D0 time point. The mean acetate concentration in the D7 cellulose-supplemented cultures was found to be 106.3 mM, compared to 75.0 mM in the D7 control replicates, which then decreased in the D14 experimental and control replicates to 89.5 mM and 56.4 mM, respectively. Propionate concentrations were found to be 6.8 mM in the rumen, which increased to 39.8 mM and 17.3 mM in the D7 cellulose-enriched and unsupplemented replicates, respectively, and then slightly decreased to 31.1 mM and 14.1 mM in the D14

cultures. The rumen butyrate concentration was found to be 5.2 mM, and in D7 experimental cultures, was found to be 13.8 mM, in contrast to 9.5 mM in the D7 control cultures. Butyrate concentraions decreased in both the D14 cellulose-enriched and control cultures to 11.7 mM and 8.3 mM, respectively.

Sample 2 acetate concentrations was found to be 50.2 mM in the rumen, which increased to a mean concentration of 90.5 mM and 95.5 mM in the D7 control and experimental cultures, respectively. Concentrations then decreased to 76.5 mM and 90.7 mM in the D14 unsupplemented and cellulose-enriched cultures, respectively. Ruminal propionate concentrations were found to be 12.5 mM, which then increased to 20.3 mM in the D7 control cultures and 23.0 mM in the experimental cultures. In the D14 cultures, propionate concentrations slightly decreased in the control (18.8 mM) and cellulose-supplemented cultures (22.1 mM). Butyrate concentrations in the rumen were found to be 8.2 mM, which increased to 11.2 mM and 12.2 mM in the D7 control and experimental cultures, respectively. D14 butyrate concentrations decreased in the control cultures (10.4 mM), in contrast to the cellulose-supplemented cultures which increased (12.4 mM). Sample 2 had an initial pH of 6.46, and tended to decrease in both the control and experimental cultures, with time (Figure 2.3). The mean pH in the d7 control and experimental samples was found to be 5.82 and 5.41, respectively, whereas the d14 control and experimental samples showed pH levels of 6.0 and 5.39.

Cow 3 had an initial pH of 6.8 which decreased to an average of 6.03 and 5.4 in the d7 control and experimental samples, respectively (Figure 2.3). On d14, the mean pH for the control did not differ from that found on d7; however, pH levels slightly increased in the experimental cultures to 5.46.

Discussion

In ruminants, microorganisms that populate the rumen compartment of the gastro-intestinal tract are responsible for digesting the different components of feed, producing end products in the form of VFAs that can be absorbed and metabolized by their host (Van Soest, 1994). As a group, rumen microorganisms are very diverse, with an estimated several hundred to several thousand species (Creevey *et al.*, 2014). Since the majority of rumen microorganisms still remain to be characterized, the factors responsible for the vast complexity of this ecosystem have yet to be fully elucidated. Host-symbiont relationships, the high degree of specialization of anaerobic microorganisms, as well as the wide array of molecular components that compose ruminant diets are typically recognized as major determinants of rumen microbial composition (Kim *et al.*, 2011; Creevey *et al.*, 2014).

Among the diverse array of organic material found in the diet of a ruminant, there is considerable interest in further elucidating the mechanisms involved in the breakdown of cellulose. Indeed, not only is cellulose the most abundant component of forage-based diets, but hosts are dependent on microbial symbionts for its digestion since mammalian genomes do not encode the necessary enzymes (Jami & Mizrahi, 2012). Efforts to gain further insight have included the isolation of species as well as metagenomics analyses. While a number of cellulose utilizing rumen microbial species have been isolated, metagenomics studies have revealed that the vast majority remain to be identified or characterized (Koike *et al.*, 2003; Brulc *et al.*, 2009; Hess *et al.*, 2011; Sirohi *et al.*, 2012; Dai *et al.*, 2015). In an effort to gain further insight, we have described in this report five species-level OTUs that were enriched in rumen cultures supplemented with cellulose.

Of the OTUs identified in our study, four were assigned to *Ruminococcaceae*. Members of this family were among the most abundant in the rumen samples analyzed, ranging from 6.1%-10.3% of the total reads per sample, then increasing to 15.8%-19.3% within 7 days in the presence of cellulose. Previous studies have shown that *Ruminococcaceae* can be very highly represented in the rumen. Indeed, the meta-analysis by Kim *et al*. (2011) revealed that *Ruminococcaceae* represented approximately 13.5% of all full length bacterial rumen 16S rRNA gene sequences, corresponding to 25.8% of *Clostridia*, in publicly available databases.

OTU C10-00008 was found to correspond to *R. flavefaciens*, based on our genetic distance cutoff. It is a known cellulose-utilizing species that expresses cellulosomes, which are cell surface protein complexes that can bind to plant fiber particles and hydrolyze plant cell wall polysaccharides (Bayer *et al.*, 2008). Various strains of *R. flavefaciens* have been shown to express a broad range of carbohydrate-utilizing enzymes, including endocellulase, glucanase, exoglucanase, and cellodextrinase, making it a versatile user of plant cell wall fiber polymers (Gardner *et al.*, 1987; Howard & White, 1988; Doerner & White, 1990; Wang & Thomson, 1990; Wang *et al.*, 1993; Vercoe *et al.*, 1995a; Vercoe *et al.*, 1995b). Interestingly, its crystalline cellulose metabolizing rate has been shown to be higher than other fibrolytic rumen bacteria, indicating that this microorganism would likely thrive in fiber-rich environments (Latham *et al.*, 1978a; Latham *et al.*, 1978b; Shi & Weimer, 1992; Miron *et al.*, 1994). While OTUs related to *R. flavefaciens* have commonly been reported from ruminal diversity studies, they are typically found in relatively low abundance. For instance, Creevy *et al*. (2014) described two distinct strains of *R. flavefaciens* as part of the core rumen bacteria,

which represented 0.06% to 0.08% of the total reads on a scaled average among seven studies. These values are slightly lower, but within the same range, of the relative abundance found in the rumen sample from which it was enriched (0.21%). While *R. flavefaciens*, together with other well-characterized fibrolytic bacteria such as *Fibrobacter succinogenes* and *R. albus,* typically do not account for more than 1% to 3% of the total bacteria in the rumen, they are expected to contribute significantly in rumen function (Stevenson & Weimer, 2007; Flint *et al.*, 2008; Wallace, 2008; Russell *et al.*, 2009; Mosoni *et al.*, 2011).

Two OTUs from this taxonomic group were designated as unclassified *Ruminococcaceae*. In their meta-analysis, Kim *et al.* (2011) reported that approximately 8.7% of reads corresponded to unclassified *Ruminococcaceae*, which is within the range observed for the rumen samples collected in this study. Jami *et al*. (2013) observed a slightly lower relative abundance for unclassified *Ruminococcaceae* (5.56%) in the rumen of 2-year old Israeli Holsteins that were fed a 70% concentrate: 30% roughage diet. Evidence supporting their involvement in fiber degradation was provided by Huws *et al*. (2013), who reported an increase in the abundance of members from this family after 8 hours of *in situ* incubation with fresh perennial ryegrass. Fouts *et al*. (2012) also reported an unclassified *Ruminococcaceae* OTU as one of the most abundant rumen bacteria in cows consuming a forage diet.

Intriguingly, an uncharacterized species of the *Ruminococcaceae* genus *Ethanoligenens* was also enriched from sample 2. *Ethanoligenens harbinense* is thus far the only species identified from this genus. Originally isolated from the anaerobic sludge of molasses wastewater, it is capable of fermenting various mono-, di- and

oligosaccharides, including D-glucose and cellobiose, into ethanol, acetate, hydrogen, and carbon dioxide (Xing *et al.*, 2006). Based on these metabolic activities, co-enrichment of an uncharacterized species of *Ethanoligenens* suggests that it may benefit from the cellulose metabolizing activity of other rumen bacteria, such as OTUs LOD11-02512 and LOD10-23342 that were also enriched in sample 2. Alternatively, the uncharacterized *Ethanoligenens* identified in our study may be capable of metabolizing cellulose. Indeed, Chang *et al*. (2010) showed *E. harbinense* to be among the main species found in the second generation of a repeated-batch culture of ruminal fluid supplemented with napiergrass. Moreover, an OTU corresponding to an *Ethanoligenens*-related species was found in rumen fluid supplemented with wheat straw (Li *et al.*, 2017). Identifying this organism in this study, as well as in other studies supplementing rumen inoculum with cellulosic biomass, suggests that it may contribute to plant fiber breakdown.

Unexpectedly, only one enriched OTU was assigned to *Prevotella*. Indeed, this genus has not only the highest number of species amongst *Bacteroidetes* genera, it is also considered to be the most predominant bacterial group in the rumen (Stevenson & Weimer, 2007; Jami & Mizrahi, 2012). It was the most frequent genus found by Kim *et al*. (2011), accounting for 11.1% of all bacterial sequences, including 41.5% of *Bacteroidetes* sequences. Several species belonging to the genus *Prevotella* have been previously described to have cellulolytic capabilities, including *P. bryantii*, *P. rumincola,* and *P. albensis*, and are regularly found in rumen diversity studies (Stevenson & Weimer, 2007; Purushe *et al.*, 2010). The importance of this group in fiber degradation in the rumen is supported by the number of enzymes identified in this genus, including

endocellulase, endogluconase, and exogluconase (Vercoe & Gregg, 1992; Wulff-Strobel & Wilson, 1995; Krause *et al.*, 2003). Further support for the contribution of *Prevotella* species in fiber degradation was shown by a metagenomics study by Jose *et al*. (2017), where *Prevotella* was reported to account for more than 36% of the glycoside hydrolases and other carbohydrate-metabolizing enzymes identified in animals fed finger millet straw.

While members of the *Prevotellaceae* family were found to be the most abundant group in the rumen samples analyzed in this study, their relative abundance decreased after 7 days in the presence of cellulose as the only exogenous substrate provided. Since certain valid *Prevotella* species have been found to metabolize starch and proteins rather than cellulose (Griswold & Mackie, 1997; McSweeney & Mackie, 2012), perhaps this group was involved in metabolizing other components of the diet for the rumen samples analyzed in this study.

Conclusion

Overall, supplementation with cellulose allowed the enrichment of one to three OTUs from bovine rumen fluid sampled from different individual hosts. While further investigation is needed to determine their genetic and metabolic potential, the ability of these OTUs to thrive in the presence of cellulose suggests their involvement in cellulose degradation. Moreover, enrichment of one known cellulolytic organism supports the approach used in this study. Identification of such bacteria could allow for a greater understanding of ruminal cellulose metabolism, with the potential to offer new insight on strategies to enhance efficiency of cellulose use in ruminant livestock. Further insight

into their potential could be obtained through functional-based studies, metagenomics, or both, which would allow us to explore their genetic and metabolic capabilities.

Literature Cited

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Bayer EA, Lamed R, White BA & Flint HJ (2008) From cellulosomes to cellulosomics. *Chemical Record (New York, NY)* **8**: 364-377.

Brulc JM, Antonopoulos DA, Miller ME*, et al.* (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences of the USA* **106**: 1948-1953.

Chang J-J, Lin J-J, Ho C-Y, Chin W-C & Huang C-C (2010) Establishment of rumen-mimic bacterial consortia: A functional union for bio-hydrogen production from cellulosic bioresource. *International Journal of Hydrogen Energy* **35**: 13399-13406.

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

Dai X, Tian Y, Li J*, et al.* (2015) Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in cow rumen. *Applied and Environmental Microbiology* **81**: 1375-1386.

Doerner KC & White BA (1990) Assessment of the endo-1,4-beta-glucanase components of Ruminococcus flavefaciens FD-1. *Applied and Environmental Microbiology* **56**: 1844-1850.

Edwards U, Rogall T, Bloecker H, Emde M & Boettger EC (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research* **17**: 7843–7854.

Flint HJ, Bayer EA, Rincon MT, Lamed R & White BA (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Reviews Microbiology* **6**: 121-131.

Fouts DE, Szpakowski S, Purushe J, Torralba M, Waterman RC, MacNeil MD, Alexander LJ & Nelson KE (2012) Next generation sequencing to define prokaryotic and fungal diversity in the bovine rumen. *PloS One* **7**: e48289.

Gardner RM, Doerner KC & White BA (1987) Purification and characterization of an exo-beta-1,4-glucanase from Ruminococcus flavefaciens FD-1. *Journal of Bacteriology* **169**: 4581-4588.

Griswold KE & Mackie RI (1997) Degradation of protein and utilization of the hydrolytic products by a predominant ruminal bacterium, Prevotella ruminicola B1(4). *Journal of Dairy Science* **80**: 167-175.

Hess M, Sczyrba A, Egan R*, et al.* (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science (New York, NY)* **331**: 463–467.

Howard GT & White BA (1988) Molecular Cloning and Expression of Cellulase Genes from Ruminococcus albus 8 in Escherichia coli Bacteriophage lambda. *Applied and Environmental Microbiology* **54**: 1752-1755.

Huws SA, Mayorga OL, Theodorou MK, Onime LA, Kim EJ, Cookson AH, Newbold CJ & Kingston-Smith AH (2013) Successional colonization of perennial ryegrass by rumen bacteria. *Letters in Applied Microbiology* **56**: 186-196.

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

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

Johnson DR, Goldschmidt F, Lilja EE & Ackermann M (2012) Metabolic specialization and the assembly of microbial communities. *ISME J* **6**: 1985-1991.

Jose VL, Appoothy T, More RP & Arun AS (2017) Metagenomic insights into the rumen microbial fibrolytic enzymes in Indian crossbred cattle fed finger millet straw. *AMB Express* **7**: 13.

Kim M, Morrison M & Yu Z (2011) Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiological Ecology* **76**: 49–63.

Koike S, Yoshitani S, Kobayashi Y & Tanaka K (2003) Phylogenetic analysis of fiber-associated rumen bacterial community and PCR detection of uncultured bacteria. *FEMS Microbiology Letters* **229**: 23-30.

Krause DO, Denman SE, Mackie RI, Morrison M, Rae AL, Attwood GT & McSweeney CS (2003) Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. *FEMS Microbiology Reviews* **27**: 663-693.

Lane D, Pace B, Olsen GJ, Stahl DA, Sogin ML & Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the USA* **82**: 6955-6959.

Latham MJ, Brooker BE, Pettipher GL & Harris PJ (1978a) Ruminococcus flavefaciens Cell Coat and Adhesion to Cotton Cellulose and to Cell Walls in Leaves of Perennial Ryegrass (Lolium perenne). *Applied and Environmental Microbiology* **35**: 156-165.

Latham MJ, Brooker BE, Pettipher GL & Harris PJ (1978b) Adhesion of Bacteroides succinogenes in pure culture and in the presence of Ruminococcus flavefaciens to cell walls in leaves of perennial ryegrass (Lolium perenne). *Applied and Environmental Microbiology* **35**: 1166-1173.

Ley RE, Hamady M, Lozupone C*, et al.* (2008) Evolution of mammals and their gut microbes. *Science (New York, NY)* **320**: 1647–1651.

Li K, Zhu H, Zhang Y & Zhang H (2017) Characterization of the microbial communities in rumen fluid inoculated reactors for the biogas digestion of wheat straw. *Sustainability* **9**: 243.

Macfarlane GT & Macfarlane S (1993) Factors affecting fermentation reactions in the large bowel. *The Proceedings of the Nutrition Society* **52**: 367-373.

McSweeney C & Mackie R (2012) Micro-organisms and ruminant digestion: state of knowledge, trends and future prospects. Background study paper No. 61. Food Agrie. Org. United Nations, Rome, Italy.

Miron J, Duncan SH & Stewart CS (1994) Interactions between rumen bacterial strains during the degradation and utilization of the monosaccharides of barley straw cell-walls. *Journal of Applied Bacteriology* **76**: 282-287.

Mosoni P, Martin C, Forano E & Morgavi DP (2011) Long-term defaunation increases the abundance of cellulolytic ruminococci and methanogens but does not affect

the bacterial and methanogen diversity in the rumen of sheep. *Journal of Animal Science* **89**: 783-791.

Naas AE, Mackenzie AK, Mravec J, Schückel J, Willats WG, Eijsink VG & Pope PB (2014) Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio* **5**: e01401.

Purushe J, Fouts DE, Morrison M, White BA, Mackie RI, Coutinho PM, Henrissat B & Nelson KE (2010) Comparative genome analysis of Prevotella ruminicola and Prevotella bryantii: insights into their environmental niche. *Microbial Ecology* **60**: 721-729.

Russell JB, Muck RE & Weimer PJ (2009) Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. *FEMS Microbiology Ecology* **67**: 183-197.

Schloss PD, Westcott SL, Ryabin T*, et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**: 7537-7541.

Shi Y & Weimer PJ (1992) Response surface analysis of the effects of pH and dilution rate on Ruminococcus flavefaciens FD-1 in cellulose-fed continuous culture. *Applied and Environmental Microbiology* **58**: 2583-2591.

Sirohi SK, Singh N, Dagar SS & Puniya AK (2012) Molecular tools for deciphering the microbial community structure and diversity in rumen ecosystem. *Applied Microbiology and Biotechnology* **95**: 1135-1154.

St-Pierre B & Wright A-DG (2015) Investigation of bacterial and methanogen community composition and diversity in full-scale anaerobic manure digesters. *BAOJ Microbiology* **1**: 1.

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

Sträuber H, Schröder M & Kleinsteuber S (2012) Metabolic and microbial community dynamics during the hydrolytic and acidogenic fermentation in a leach-bed process. *Energy, Sustainability and Society* **2**: 13.

Thornton PK (2010) Livestock production: recent trends, future prospects. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **365**: 2853-2867.

Van Soest PJ (1994) Nutrional ecology of the ruminant (second edition). 476 pages.

Vercoe PE & Gregg K (1992) DNA sequence and transcription of an endoglucanase gene from Prevotella (Bacteroides) ruminicola AR20. *Molecular and General Genetics* **233**: 284-292.

Vercoe PE, Finks JL & White BA (1995a) DNA sequence and transcriptional characterization of a beta-glucanase gene (celB) from Ruminococcus flavefaciens FD-1. *Canadian Journal of Microbiology* **41**: 869-876.

Vercoe PE, Spight DH & White BA (1995b) Nucleotide sequence and transcriptional analysis of the celD beta-glucanase gene from Ruminococcus flavefaciens FD-1. *Canadian Journal of Microbiology* **41**: 27-34.

Wallace JR (2008) Gut microbiology – broad genetic diversity, yet specific metabolic niches. *Animal* **2**: 661-668.

Wang L, Hatem A, Catalyurek UV, Morrison M & Yu Z (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PloS One* **8**: e78507.

Wang Q, Garrity GM, Tiedje JM & Cole JR (2007) Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**: 5261-5267.

Wang W, Reid SJ & Thomson JA (1993) Transcriptional regulation of an endoglucanase and a cellodextrinase gene in Ruminococcus flavefaciens FD-1. *Journal of General Microbiology* **139 Pt 6**: 1219-1226.

Wang WY & Thomson JA (1990) Nucleotide sequence of the celA gene encoding a cellodextrinase of Ruminococcus flavefaciens FD-1. *Molecular & General Genetics* **222**: 265-269.

Wulff-Strobel CR & Wilson DB (1995) Cloning, sequencing, and characterization of a membrane-associated Prevotella ruminicola B(1)4 beta-glucosidase with cellodextrinase and cyanoglycosidase activities. *Journal of Bacteriology* **177**: 5884-5890.

Xing D, Ren N, Li Q, Lin M, Wang A & Zhao L (2006) Ethanoligenens harbinense gen. nov., sp. nov., isolated from molasses wastewater. *International Journal of Systematic and Evolutionary Microbiology* **56**: 755-760.

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

Yue ZB, Li WW & Yu HQ (2013) Application of rumen microorganisms for anaerobic bioconversion of lignocellulosic biomass. *Bioresource Technology* **128**: 738-744.

Figure 2.1. Taxonomy-based bacterial community profile of three rumen samples prior to cellulose enrichment. Values are expressed as relative abundance (%). A) Rumen sample 1; B) Rumen sample 2; C) Rumen sample 3; D) Color-coded legend. Other phyla category include *Proteobacteria*, *Fibrobacteres*, *Lentisphaerae, Spirochaetes, Tenericutes, Fusobacteria, Elusimicrobia, Actinomycetes,* and/or *Planctomycetes*. (uncl.= unclassified)

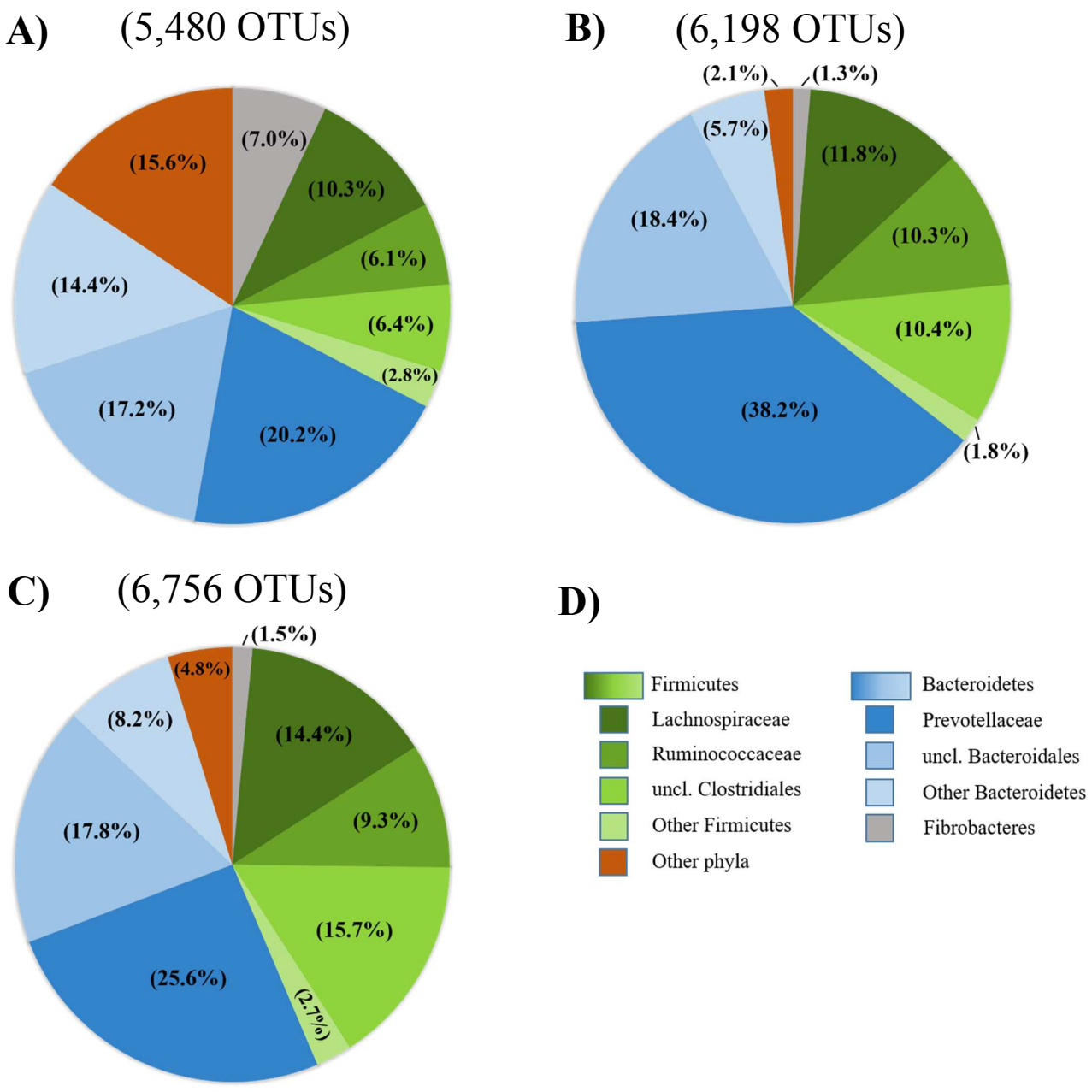

Figure 2.2. Relative abundance of five enriched bacterial species-level OTUs in the presence of cellulose. Relative abundances for uncultured rumen samples (D0) and cellulose-supplemented cultures (Cel 1, Cel 2, and Cel 3) at D7 and D14 are presented separately, while means of duplicates are shown for control cultures. Predicted taxonomic affiliations for each of the represented OTUs in A, B, and C are shown in the figure legend as determined by RDP Classifier and BLAST.

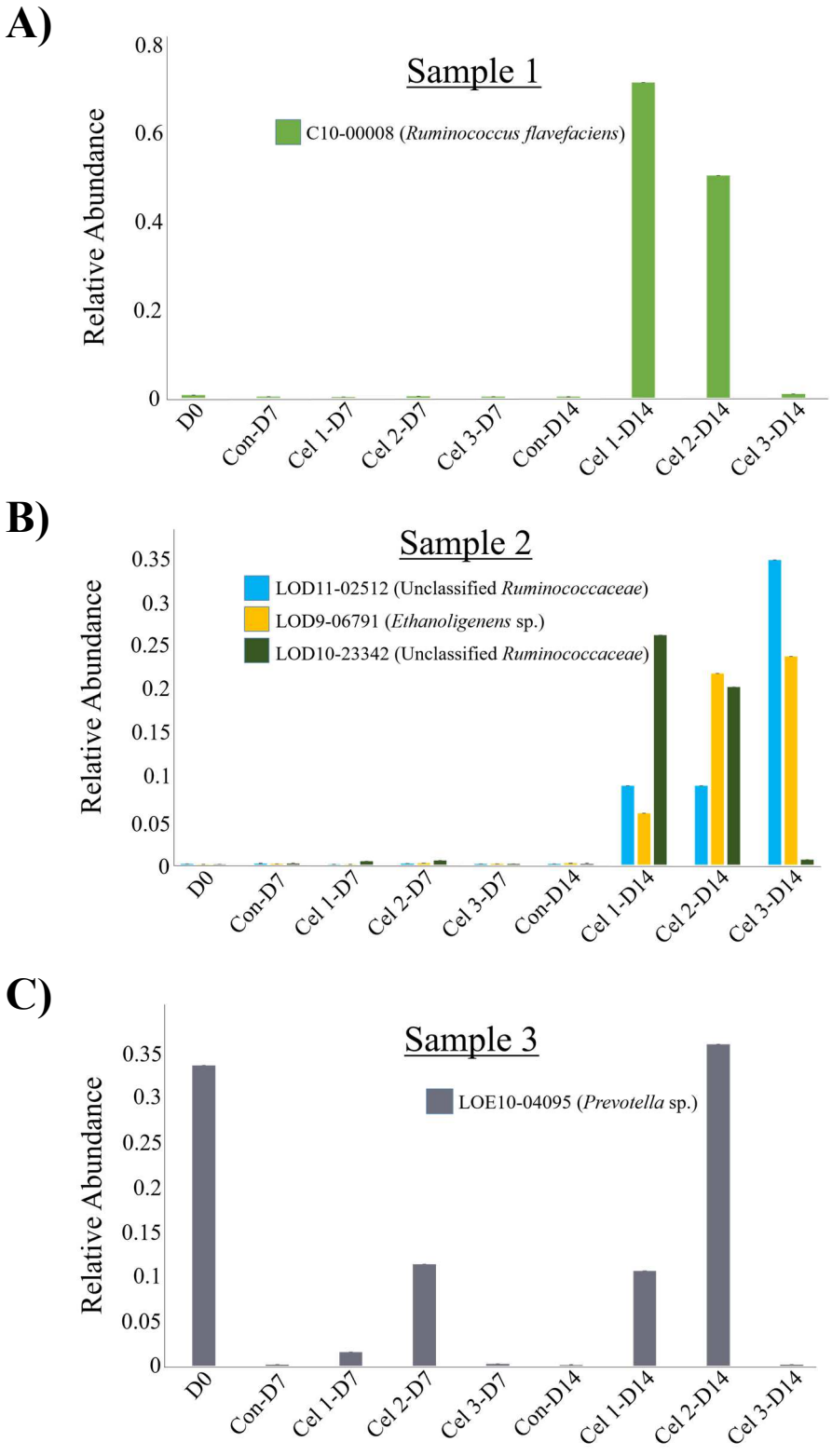

Figure 2.3. Mean pH values for cellulose cultures (triplicate) and control cultures (duplicate) in sample 2 (A) and sample 3 (B) on selected days of incubation.

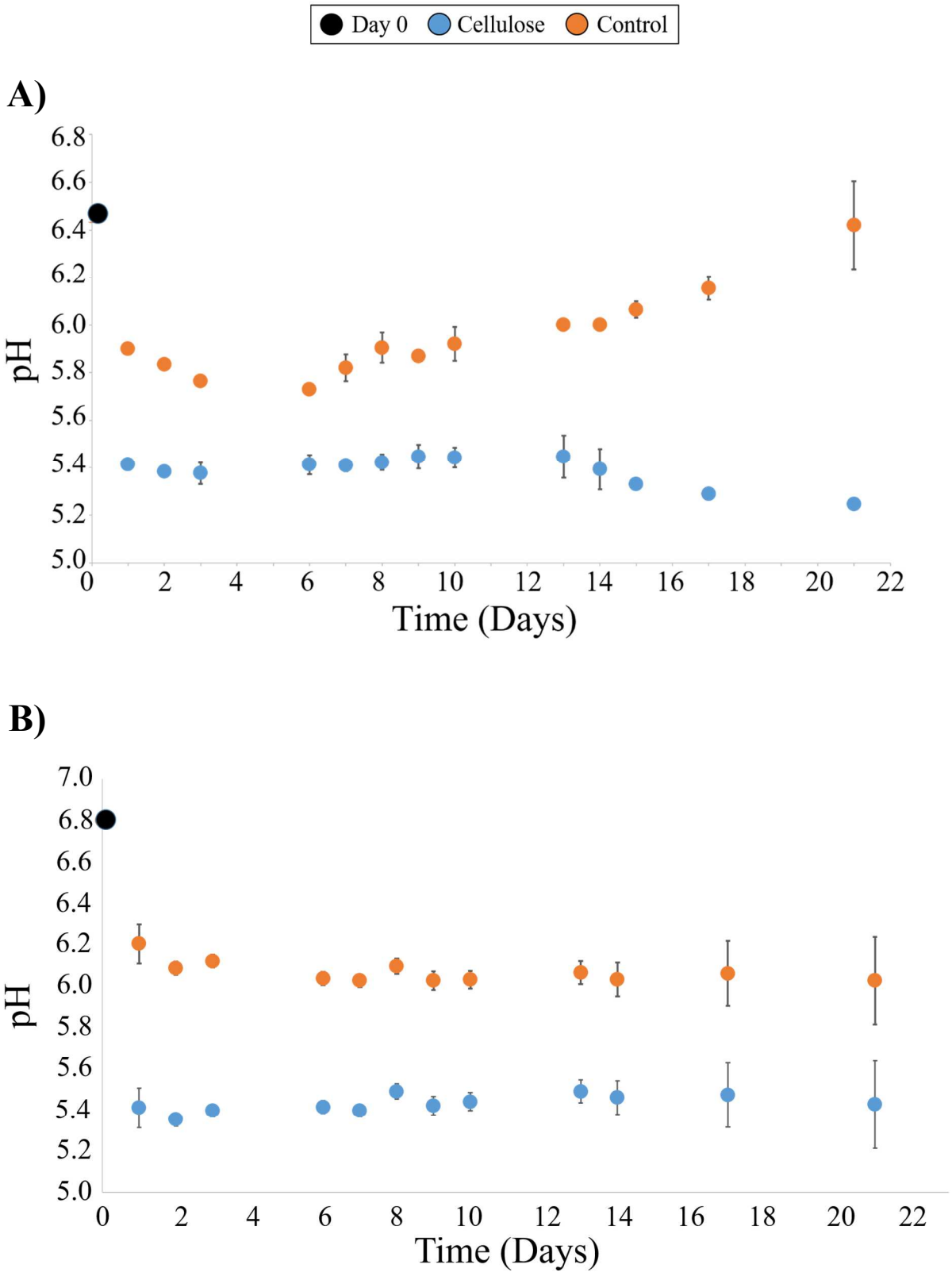

Table 2.1. Number of high-quality, non-chimeric 16S rRNA reads used to determine the bacterial composition of samples from rumen fluid and at day 7 (D7) and day 14 (D14).

	Rumen[a]	D7		D14	
		Con[b]	Cel[c]	Con[b]	Cel[c]
Sample 1	16,278	10,246-22,121	9,281-16,942	13,642-16,681	11,017-19,584
Sample 2	8,348	5,678-7,299	8,486-9,479	4,368-10,226	28,773-40,028
Sample 3	8,236	10,660-12,935	11,171-13,969	4,587-5,954	7,617-13,523

[a] Single sample
[b] Range for 2 samples
[c] Range for 3 samples

Table 2.2. Relative abundances (%) of major taxonomic groups from rumen samples (D0) and cellulose-enriched cultures (D7).

Taxonomic Affiliation		Sample 1		Sample 2		Sample 3	
		D0[a]	D7[b]	D0[a]	D7[b]	D0[a]	D7[b]
Bacteroidetes		51.8	35.3±5.3	62.3	35.7±3.9	51.6	34.3±1.8
	Prevotellaceae	20.2	11.6±1.6	38.2	20.2±1.6	25.6	22.6±1.7
	Unclassified *Bacteroidales*	17.2	10.6±6.5	18.4	10.1±1.8	17.8	7.4±1.8
	Other *Bacteroidetes*	14.4	13.1±5.9	5.7	5.4±1	8.2	4.3±0.6
Firmicutes		25.6	44.6±2.7	34.3	50.3±6	42.1	54±1.3
	Ruminococcaceae	6.1	15.8±6.1	10.3	17.5±2.7	9.3	19.3±1.3
	Lachnospiraceae	10.3	8.1±2.2	11.8	8.2±0.9	14.4	15.3±1.8
	Unclassified *Clostridiales*	6.4	14.7±4.1	10.4	17.7±1.8	15.7	14.4±0.8
	Other *Firmicutes*	2.8	6±1.5	1.8	6.9±1.3	2.7	5±1
Fibrobacteres		7	0.3±0.3	1.3	9.5±8.2	1.5	0.1±0.1
Other Bacteria		15.6	19.8±3.1	2.1	4.5±1.8	4.8	11.6±0.6

[a]Value obtained from single sample
[b]Average and standard deviation from triplicate cultures

Table 2.3. Relative abundance and RDP Classifier derived taxonomic assignment for highly represented (>1%) OTUs in rumen fluid, day 7 (D7) control (Con), and/or cellulose-supplemented (Cel) cultures in sample 1.

OTU	D0[a]	D7		Taxonomy
		Con[b]	Cel[c]	
C4-00892	0.4	0.2±0.1	7.0±2.9	*Saccharibacteria_genera_incertae_sedis*
C3-01379	0.3	0.9±0.7	4.6±3.2	unclassified_*Ruminococcaceae*
C1-00068	3.9	2.3±0.6	2.2±0.8	unclassified_Bacteria
C2-00501	0.04	6.8±7.2	0.1±0.01	*Arcobacter*
C2-00282	0.01	6.0±5.7	0.02±0.01	*Alcaligenes*
C2-05094	0.1	2.7±3.6	0.3±0.2	unclassified_Bacteria
C1-16337	1.8	0.05±0.02	2.4±0.4	*Prevotella*
C1-05580	1.3	0.5±0.5	1.8±0.6	unclassified_*Planctomycetaceae*
C2-13598	0.6	1.4±0.6	1.0±0.4	unclassified_*Ruminococcaceae*
C1-02448	1.2	1.3±1.3	0.7±0.3	unclassified_*Planctomycetaceae*
C4-02376	0.3	0.1±0.1	2.0±0.7	*Saccharibacteria_genera_incertae_sedis*
C1-20114	0.8	0.05±0.001	1.9±1.8	unclassified_*Bacteroidales*
C1-00377	4.7	0.04±0.02	0.2±0.2	*Fibrobacter*
C1-17299	0.9	0.1±0.01	1.5±1.3	unclassified_*Bacteroidales*
C1-03796	0.7	0.2±0.1	1.2±0.9	*Succiniclasticum*
C1-01207	1.7	0.2±0.1	0.8±0.5	unclassified_*Porphyromonadaceae*
C1-03174	0.1	1.5±0.8	0.3±0.01	unclassified_*Bacteroidetes*
C4-08978	0.05	0.1±0.04	1.5±0.7	unclassified_*Ruminococcaceae*
C1-03126	0.4	1.0±1.1	0.3±0.2	unclassified_*Planctomycetaceae*
C3-06973	0.3	0.2±0.1	1.1±0.8	unclassified_*Clostridiales*
C1-06399	0.1	1.5±0.6	0.3±0.2	unclassified_*Bacteroidetes*
C2-06248	0.01	1.5±1.6	0.5±0.4	*Comamonas*
C4-00066	1.2	0.01±0.001	0.7±0.4	*Prevotella*
C2-06434	0.02	1.1±0.5	0.04±0.02	*Victivallis*

[a] Value obtained from single sample
[b] Average and standard deviation from duplicate cultures
[c] Average and standard deviation from triplicate cultures

Table 2.4. Relative abundance and RDP Classifier derived taxonomic assignment for highly represented (>1%) OTUs in rumen fluid, day 7 (D7) control (Con), and/or cellulose-supplemented (Cel) cultures in sample 2.

	D0[a]	D7		
OTU		Con[b]	Cel[c]	Taxonomy
LOD5-02224	0.7	0.01±0.01	7.7±6.9	*Fibrobacter*
LOD4-01417	0.1	0.01±0.02	1.6±0.4	*Prevotella*

[a] Value obtained from single sample
[b] Average and standard deviation from duplicate cultures
[c] Average and standard deviation from triplicate cultures

Table 2.5. Relative abundance and RDP Classifier derived taxonomic assignment for highly represented (>1%) OTUs in rumen fluid, day 7 (D7) control (Con), and/or cellulose-supplemented (Cel) cultures in sample 3.

OTU	D0[a]	D7		Taxonomy
		Con[b]	Cel[c]	
LOE10-04095	33.6	0.1±0.03	4.4±6.1	*Prevotella*
LOE4-02289	0.1	0.1±0.02	4.2±6.0	*Prevotella*
LOE2-14324	0.2	0.4±0.3	2.5±0.2	unclassified_*Alphaproteobacteria*
LOE2-14329	0.2	0.3±0.3	1.0±0.1	unclassified_*Ruminococcaceae*
LOE3-00201	1.0	0.6±0.2	0.6±0.1	*SR1_genera_incertae_sedis*
LOE4-22899	0.01	0.03±0.002	1.1±1.6	unclassified_*Ruminococcaceae*
LOE2-01576	0.2	1.3±0.8	0.1±0.02	unclassified_*Bacteroidales*

[a] Value obtained from single sample
[b] Average and standard deviation from duplicate cultures
[c] Average and standard deviation from triplicate cultures

Table 2.6. Rumen and culture VFA concentrations from sample 1 (A) and sample 2 (B).

A)

	Acetate		Propionate		Butyrate	
	mM	**SD**	**mM**	**SD**	**mM**	**SD**
D0[a]	17.6*	0	6.8	0	5.2	0
Con D7[b]	75.0	7.4	17.3	0.9	9.5	0.2
Cel D7[c]	106.3	3.5	39.8	0.6	13.8	0.4
Con D14[b]	56.4	6.6	14.1	1.2	8.3	0.8
Cel D14[c]	89.5	9.8	31.1	5.9	11.7	1.5

B)

	Acetate		Propionate		Butyrate	
	mM	**SD**	**mM**	**SD**	**mM**	**SD**
D0[a]	50.2	0	12.5	0	8.2	0
Con D7[b]	90.5	11.1	20.3	2.7	11.2	1.0
Cel D7[c]	95.5	7.6	23.0	0.8	12.2	0.3
Con D14[b]	76.5	9.9	18.8	2.5	10.4	0.8
Cel D14[c]	90.7	10.9	22.1	2.6	12.4	0.9

* Value significantly lower than normal ruminal acetate concentrations
[a] Value obtained from single sample
[b] Average and standard deviation from duplicate cultures
[c] Average and standard deviation from triplicate cultures

Chapter 3

METAGENOMIC ANALYSIS OF TWO CANDIDATE CELLULOSE UTILIZING BACTERIA FROM THE RUMEN OF BEEF CATTLE

Abstract

The microbial-driven process of converting cellulosic biomass into utilizable energy is a defining feature of ruminant animals. Due to the complexity of rumen microbial communities, the vast majority of ruminal microorganisms remain uncharacterized. In order to gain further insight, selection-based batch culturing from bovine ruminal fluid in combination with metagenomics was used to identify previously uncharacterized rumen bacteria capable of metabolizing cellulose. 16S rRNA-based population analysis was used to identify bacteria enriched from ruminal fluid within 14 days of culturing in the presence of cellulose. Two distinct species-level operational taxonomic units (OTUs) were enriched. One OTU corresponding to an uncultured species of *Rummeliibacillus* showed a 291-fold average increase compared to the original rumen inoculum, and represented 30.7% of reads in cultures supplemented with cellulose. The other OTU was an uncultured species of *Prevotella* sp. and had increased by 46-fold after 14 days in the presence of cellulose, and represented on average 14.1% of the reads. Metagenome sequencing and analysis revealed genes with predicted cellulolytic capabilities (cellulase, endoglucanase, and beta-glucanase) in the *Rummeliibacillus* and *Prevotella*-related organism, as well as genes predicted to function in cellulose binding. Functions involved in amino acid acquisition such as protease activity, amino acid transport and glutamate synthase, were also found in both OTUs. The identification and characterization of novel candidate cellulolytic species of ruminal bacteria provides additional insight towards a better understanding of ruminal cellulose metabolism.

Introduction

Cellulose, the main component of plant cell walls, represents one of the most abundant organic polymers and reserve of monosaccharides on earth (Naas *et al.*, 2014). As a result, herbivory has successfully evolved in certain animal groups to utilize this abundant source of energy. However, since mammalian genomes do not encode for cellulose-metabolizing enzymes, herbivores typically rely on commensal gastrointestinal microorganisms to breakdown plant-cell wall polysaccharides (Van Soest, 1994; McSweeney & Mackie, 2012; Yue *et al.*, 2013).

Due to their complex molecular structures, cellulose and other components of plant fibers represent recalcitrant substrates even for microorganisms that can metabolize them. Consequently, mammalian herbivory has successfully evolved two main strategies to maximize the digestion efficiency of these complex substrates. In hindgut fermenters, the distal segments of the gastrointestinal tract are the most developed, resulting in increased intestinal volume to favor retention time and maintenance of high microbial numbers (Macfarlane & Macfarlane, 1993). In contrast, ruminants have developed a compartmentalized stomach, of which the largest segment, known as the rumen, hosts the microbial symbionts responsible for digesting the feed consumed by the animal (Van Soest, 1994). Based on the number of ruminant species that have evolved and the diverse array of habitats they occupy as a group, foregut fermentation represents a very successful strategy for herbivory. Thanks to their ability to transform inedible plant biomass into protein-rich products, ruminants have played a central role throughout human history, and, to this day, are expected to contribute in meeting the demands of a rapidly growing and urbanizing global population (Thornton, 2010).

The rumen microbial ecosystem consists of a diverse array of members from the three domains of life: i.e. bacteria, archaea, and eukarya (protozoa and fungi). Because they thrive in anoxic conditions, rumen microorganisms tend to have specialized metabolic functions, and thus rely on other community members to provide them with nutrients that they are unable to synthesize. As a result of this co-dependence, rumen microorganisms form complex trophic relationships (Johnson *et al.*, 2012). Based on the metabolic properties of currently known rumen microorganisms, not all members of the rumen microbial communities are involved in cellulose metabolism (Varel & Dehority, 1989; Krause *et al.*, 2003). While protozoa and fungi are major contributors, bacteria are the most investigated subgroup as they represent the most diverse group in terms of genetic and metabolic functions, and are also found at the highest cell density in the rumen (Van Soest, 1994). A number of isolates from major bacterial phyla such as *Bacteroidetes*, *Firmicutes*, and *Fibrobacteres* have been cultured, which include well characterized species such as *Prevotella ruminicola*, *Ruminococcus albus*, and *Fibrobacter succinogenes*, respectively. While these cultured isolates are thought to play an important role in normal ruminal function, they may represent only 6% of rumen cellulose utilizing bacterial species (Russell *et al.*, 2009; Creevey *et al.*, 2014).

Genomic sequencing of cultured isolates has revealed an abundance of genes involved in cellulose catabolism. For example, *R. flavefaciens* FD-1 has been shown to have exoglucanase activity, a cellodextrinase, and at least four glucanases (*celB, celC, celD, celE*) (Gardner *et al.*, 1987; Howard & White, 1988; Doerner & White, 1990; Wang & Thomson, 1990; Wang *et al.*, 1993; Vercoe *et al.*, 1995a; Vercoe *et al.*, 1995b). In addition, *R. flavefaciens* strain 17 also possesses endoglucanases from glycoside

hydrolase (GH) families 5 and 44 (Cunningham *et al.*, 1991; Rincon *et al.*, 2001). Sequencing of the *R. albus* genome has also revealed a variety of fibrolytic genes including seven distinct endoglucanases, a cellobiosidase, and two β-glucosidases (Forsberg *et al.*, 1997; Miyagi *et al.*, 1998). *R. albus* strain F-40 has also been shown to encode for a cellulosome complex (Ohara *et al.*, 2000), which are cell surface protein complexes that can bind to plant fiber particles and hydrolyze plant cell wall polysaccharides (Bayer *et al.*, 2008). Indeed a number of rumen metagenomics studies have reported an abundance of previously unidentified carbohydrate-active genes, suggesting a wealth of uncharacterized bacteria involved in cellulose catabolism that have yet to be isolated (Brulc *et al.*, 2009; Hess *et al.*; 2011, Wang *et al.*, 2013). In this context, we present in this report a metagenomic analysis of two candidate cellulose metabolizing bacteria that were identified through enrichment from rumen fluid. These correspond to novel species that have previously not been described.

Materials and Methods

Rumen fluid collection and culturing

All animal procedures were approved by the Institutional Animal Care and Use Committee at South Dakota State University. Samples were collected on December 2015 from a fistulated Simental × Angus cow maintained at the South Dakota State University Cow-Calf Education and Research Facility. This individual is part of a herd whose diet consisted primarily of pasture hay or haylage. For the culture experiment, 12 L of fresh rumen fluid were collected by manually squeezing the liquid fraction from solid digesta, and the samples were stored in insulated containers during transit. For culture setup, which took place within one hour of sample collection, rumen fluid stored in separate

containers during transit were combined, mixed, and distributed amongst five bioreactors (Chemglass) with a volume of approximately 2.2 L for each culture and 0.8 L for biogas (headspace). Cellulose (20g/L, Sigma cat# C8002) was added to three replicate cultures, while the remaining two bioreactors were not supplemented with any substrate. Cultures were maintained at a temperature of 38°C, with stirring at 150rpm. Each bioreactor was equipped with a flexible plastic tube to allow for the release of excess biogas and prevent pressure build-up.

Microbial DNA extraction and PCR amplification of the 16S rRNA gene

Microbial DNA was isolated from rumen and culture samples using the repeated bead beating and column method, as described by Yu and Morrison (2004). The V1-V3 region of bacterial 16S rRNA gene sequences were PCR-amplified using the 27F forward (Edwards *et al.*, 1989) and 519R reverse primer pair (Lane *et al.*, 1985). PCR reactions were performed with the Phusion *Taq* DNA polymerase (Thermo Scientific) under the following conditions: hot start (4 min, 98 ^{o}C), followed by 35 cycles of denaturation (10s, 98 ^{o}C), annealing (30s, 50 ^{o}C) and extension (30 s, 72 ^{o}C), then ending with a final extension period (10 min, 72 ^{o}C). PCR products were separated by agarose gel electrophoresis, and amplicons of the expected size (~500bp) were excised for gel purification using the QiaexII Gel extraction kit (Qiagen). For each sample, at least 400 ng of amplified DNA were submitted to Molecular Research DNA Lab (MRDNA) (Shallowater, TX, USA) for sequencing with the Illumina MiSeq 2x300 platform to generate overlapping paired end reads.

Computational analysis of PCR-generated 16S rRNA amplicon sequences

Raw bacterial 16S rRNA gene sequences of V1-V3 amplicons were assembled as contigs from overlapping MiSeq 2x300 paired-end reads from the same flow cell clusters. Reads were then selected using custom Perl scripts to meet the following criteria: presence of both 27F (forward) and 519R (reverse) primer nucleotide sequences, length between 400 and 580nt, and a minimal quality threshold of no more than five nucleotides with a Phred quality score lower than 15. Using custom Perl scripts, these high quality reads were then aligned, and clustered into operational taxonomic units (OTUs) at a species-level genetic distance cutoff of 5% sequence dissimilarity (St-Pierre & Wright, 2015).

OTUs were then screened for sequence artifacts using three separate methods. Chimeric sequences were first identified with the chimera.uchime command from the MOTHUR open source software package (Schloss *et al.*, 2009). Secondly, OTUs with fewer than 500 reads that were not assigned to a known bacterial phylum by RDP Classifier ("unclassified bacteria" at an 80% bootstrap cutoff) (Wang *et al.*, 2007) were designated as artifacts. Finally, the integrity of the 5′ and 3′ ends of OTUs was tested using a database alignment search based approach. When compared to their closest match of equal or longer sequence length from the NCBI nt database found by BLAST (Altschul *et al.*, 1997), OTUs with more than five nucleotides missing from the 5′ or 3′ end of their respective alignments were designated as artifacts. After removal of chimeras and artifacts, taxonomic assignment of valid OTUs was performed using a combination of RDP Classifier and BLAST. The List of Prokaryotic Names with Standing in Nomenclature (LPSN) was also consulted for information on valid species belonging to various taxa of interest.

Metagenome assembly and analysis

Genomic DNA from D14 culture #3, with the highest enrichment for both OTUs of interest, was used for metagenomic analysis. Random or "shotgun" sequencing was performed using the Illumina MiSeq 2x250 (MRDNA, Shallowater, TX). Raw sequenced reads greater than 200bp in length were used in this analysis, and *de novo* assembly of the sequences was performed using ABySS (Simpson *et al.*, 2009) at a kmer length of 64bp. To ensure analysis of contigs corresponding to only the OTUs of interest, a minimal length of 5,000 nt was used as threshold for downstream analysis. Potential protein-encoding open reading frames (ORFs) within the assembled contigs were identified using the online tool RAST (Aziz *et al.*, 2008). The taxonomic affiliation of each contig was assigned using BLASTp, which was used to distinguish between contigs belonging to each OTU. Gene annotation for ORFs were performed using a combination of RAST and BLAST, while pathway information for particular ORFs of interest was found using KEGG (Kanehisa & Goto, 2000).

Results

Bacterial community structure in the uncultured rumen sample used as inoculum

Prior to selection with cellulose, the composition of the bacterial population from the rumen inoculum was analyzed (Figure 3.1). A total of 9,045 species-level OTUs were identified, with *Bacteroidetes* being the most abundant phylum (56.6%), of which *Prevotellaceae* were the highest represented family (36.1%). *Firmicutes* was the second most abundant phylum, having a relative abundance of 20.3%, with unassigned or unclassified *Clostridiales* being the most abundant group within this phyla (6.2%).

Lachnospiraceae (5.9%) and *Ruminococcaceae* (5.7%) were the predominant families within *Firmicutes*. Other well represented phyla in the rumen sample were *Fibrobacteres* (3.7%) and the candidate phylum SR1 (6.1%).

Bacterial communities were further characterized through identification of well represented OTUs, defined as representing at least 1% of sequences (Table 3.1). Seven OTUs were found to meet this criteria. The OTU with highest abundance was assigned to the candidate phylum SR1, with a relative abundance of 4.5%. Of the remaining OTUs, two belonged to *Planctomycetes*, while *Bacteroidetes*, *Fibrobacteres*, the candidate phylum SR1, and an unclassified or unassigned phylum each were represented by one OTU.

Identification of candidate cellulose utilizing rumen bacteria using batch culture enrichment

After 14 days in cultures supplemented with cellulose, two OTUs were found in greater abundance (Figure 3.2). OTU A16-00027, increased from 0.1% in the original rumen sample to 43.5%, 41.6%, and 7.0% in the triplicate cultures supplemented with cellulose, representing a 87 to 543 fold increase. In comparison, control cultures on d14 showed a mean relative abundance of 0.2% for this OTU. The closest relatives to this OTU (sequence identity of 99%) were found to be an uncultured bacterium isolated from multiple environmental sources including marine sediment (GU368058.1, GU368037.1), anaerobic sludge from a methanogenic bioreactor (HM107031.1), and heavy metal contaminated soils (DQ154284.1) (Kourtev *et al.*, 2006). *Rummeliibacillus stabekisii* (AB669588.1) was identified as its closest valid relative (99% sequence identity) (Kanno *et al.*, 2013).

OTU A16-00103 also showed an increase in the presence of cellulose, increasing from a relative abundance of 0.3% in the rumen sample to 26.3%, 9.6%, and 6.3% in the three experimental replicates on D14 (Figure 3.2). This OTU was in contrast found in low abundance in D14 cultures not supplemented with cellulose (0.1%). It was assigned to the genus *Prevotella,* and was most closely related (99% sequence identity) to uncultured bacteria isolated from bovine rumen (AB730679.1) (Nyonyo *et al.*, 2013), the rumen of cattle fed cashew nut (AB616528.1, AB616459.1, AB616404.1, AB616507.1), the rumen of cattle fed a high grain diet (HM104936.1) (Fernando *et al.*, 2010), and the rumen of buffalo (JX003974.1, JX003973.1, JX003913.1). The closest valid relative (91% sequence identity) for this OTU was found to be *P. melaninogenica* (LT684184.1).

***de novo* assembly of metagenomic DNA sequences**

A total of 366,734 contigs were generated from ABySS. Only contigs greater than 5,000 nt in length were used for further analysis, in order to focus on genome fragments with the potential for multiple ORFs to allow operon construction, as well as the higher likelihood that larger contigs would correspond to enriched organisms within the sample of interest. Of the 511 contigs that met those criteria, 322 were assigned to the enriched *Rummeliibacillus*, 174 corresponded to the enriched *Prevotella*, while the remaining 15 contigs more likely belonged to other bacterial phyla. Prediction of amino acid sequences revealed 4,988 ORFs from the *Rummeliibacillus* contigs compared to 1,070 ORFs for the *Prevotella* of interest.

Using the online tool RAST, ORFs were annotated for functional assignment using a subsystem approach, with each subsystem representing a particular metabolic function within an organism of particular interest to this study. ORFs assigned to

carbohydrate metabolism represented 4.3% and 3.5% of predicted protein sequences amongst the contigs of the uncultured *Rummeliibacillus* and *Prevotella*, respectively (Tables 3.2 & 3.3). In comparison, most ORFs from either organism could not be assigned to a particular metabolic function.

Characterization of carbohydrate-utilizing genes in the partial genomes of enriched organisms

To widen the scope for identification of genes involved in fiber degradation, ORFs were annotated using BLAST and screened by a keyword search for carbohydrate-active related enzymes. The enriched OTU corresponding to a novel *Rummeliibacillus* species was predicted to possess eleven ORFs involved in cellulose metabolism, including endoglucanase, beta-glucanase, beta-glucosidase, and other glycoside hydrolases with known cellulolytic capabilities (Table 3.4). Thirteen other ORFs were found to correspond to glycoside hydrolases, of which four ORFs pertained to starch metabolism, while the remaining nine were not associated with starch or cellulolytic capabilities or were not assigned to a GH family. Also found was an ORF related to carbohydrate binding, four ORFs involved in monosaccharide release, including cellobiose phosphorylase and aryl-phospho-beta-D-glucosidase.

The *Prevotella*-related enriched OTU was found to have an ORF corresponding to a cellulase, as well as ten other ORFs shown to encode GHs, while one ORF encoded glycogen synthase (Table 3.5). Other ORFs of interest related to carbohydrate metabolism included twelve ORFs with carbohydrate binding activity, six ORFs that encoded for enzymes that release monosaccahrides, and two ORFs involved in glycan degradation.

Identification of glycogen synthesis pathway enzymes in both enriched organisms

Another pathway of particular interest were the enzymes involved in polysaccharide metabolism for the two enriched organisms found in this experiment. Indeed, the enriched *Rummeliibacillus* organism was shown to possess a variety of ORFs predicted to be involved in glycogen synthesis, including proteins such as glycogen phosphorylase, glycogen branching protein, glycogen biosynthesis protein, and 4-alpha-glucan branching enzyme (Figure 3.3). This organism was predicted to convert extracellular cellulose into cellodextrin, then subsequently into cellobiose using endoglucanase (EC: 3.2.1.4). Cellobiose can then be phosphorylated into cellobiose-6′P for uptake into the cell, and converted to D-glucose by 6-phospho-beta-glucosidase (EC: 3.2.1.86) for use as an energy source. Phosphorylation of D-glucose by glucokinase (EC: 2.7.1.2) results in D-glucose-6′P, which is transformed into α-D-glucose-1′-P by phosphoglucomutase (EC: 5.4.2.2) by transfer of a phosphate group. ADP-glucose is then created by glucose-1′-phosphate adenylyltransferase (EC: 2.7.7.27), to produce amylose from the conversion of ADP-glucose by starch synthase (EC: 2.4.1.21). Amylose is then branched into glycogen by 1,4-alpha-glucan branching enzyme (EC: 2.4.1.18). Two mechanisms are predicted to utilize glycogen for energy: 1) glycogen is catabolized into α-D-glucose-1′-P by glycogen phosphorylase (EC: 2.4.1.1) which can then be converted back to D-glucose-6’P by phosphoglucomutase, or 2) glycogen can be hydrolyzed into maltose/dextrin by alpha-amylase (EC: 3.2.1.1), then subsequently broken down to D-glucose by maltose-glucoamylase (EC: 3.2.1.20) or 4-alpha-glucanotransferase (EC: 2.4.1.25).

Identification of protein-utilizing genes in enriched rumen OTUs

To identify genes involved in protein catabolism, ORFs were annotated using BLAST and screened by a keyword search for protein-degrading enzymes. A number of proteases/peptidases were identified for both enriched organisms. A variety of peptidases/proteases were identified in the *Rummeliibacillus* enriched organism from a number of different peptidase families including endopeptidases, amino acid-specific proteases, zinc dependent peptidases/protease, and aminopeptidases (Table 3.6). Glutamate synthase was also identified in this organism.

The *Prevotella* enriched organism possessed peptidases/proteases including C-terminal processing peptidase-3, aminopeptidase C, peptidase C10, M16, M28, S10, S41, and U32, ion protease, as well as a rhomboid family intramembrane serine protease (Table 3.7). Enzymes associated with glutamate synthesis were also identified.

Discussion

In ruminants, rumen microorganisms are responsible for the digestion of feed, which produces end products such as volatile fatty acids that are subsequently absorbed and metabolized by their host, which contribute 60-80% of their daily energy requirements (Van Soest, 1994). Advances made in next-generation sequencing have greatly contributed to elucidate the vast diversity found in the rumen ecosystem, with an estimated several hundred to several thousand species of microorganisms. Since the majority of rumen microbial species have yet to be characterized, the factors responsible for this complexity are not fully understood. Factors typically regarded as major determinants of rumen microbial composition include host-symbiont relationships, the high degree of specialization seen by anaerobic microorganisms, and the molecular nature of substrates that compose ruminant diets (Kim *et al.*, 2011; Creevey *et al.*, 2014).

Among the different substrates found in ruminant diets, cellulose, and mechanisms involved in its metabolism, have been of considerable interest due to its high abundance in forage-based diets. Since mammalian genomes do not encode for cellulolytic genes, the breakdown of cellulose in ruminants is dependent on microorganisms residing in the rumen (Jami & Mizrahi, 2012). While a number of cellulose utilizing rumen microbial species have been isolated, metagenomics studies have revealed that the vast majority remain to be identified or characterized (Koike *et al.*, 2003; Brulc *et al.*, 2009; Hess *et al.*, 2011; Sirohi *et al.*, 2012; Dai *et al.*, 2015).

In an effort to gain further insight, we have described in this report a metagenomic analysis of a rumen culture supplemented with cellulose, in which two species-level OTUs were enriched. The two OTUs enriched in the presence of cellulose in this study belonged to two distinct bacterial phyla: *Bacteroidetes* and *Firmicutes*. These phyla have been shown to be dominant in most, if not all, rumen microbial diversity studies. OTU A16-00027 was found to belong to the genus *Rummeliibacillus*, which have so far been described as gram□positive, facultative aerobic *Firmicutes* bacteria (Her & Kim, 2013). To our knowledge, a *Rummeliibacillus* species has not been identified in the rumen of cattle. However, its importance in the gut has been supported in previous research where humans were fed commercial soya milk; *Rummeliibacillus* was among the main genera to respond to these dietary changes (Ma *et al.*, 2017). This genus is typically found in soil and sediment environments as described by da Mota *et al*. (2016), where *R. stabekisii* Strain PP9 was isolated from Antarctic soil, and by Her and Kim (2013), where *R. suwonensis* was isolated from soil collected in a mountain area of South Korea. Interestingly, *R. stabekisii* has been shown to utilize and produce acid from

D-ribose, D-xylose, D-galactose, D-glucose, D-lactose, D-mannose, D-mannitol and cellobiose (Vaishampayan *et al.*, 2009).

Rummeliibacillus is a genus of the class *Bacilli*. Piao *et al*. (2014) found members of *Bacilli* to account for over 63% of the total ruminal bacterial population in a Friesian cow fed a mixed diet containing 60% fiber, suggesting this class's importance in fiber degradation. Other ruminal microbial diversity studies have found bacteria belonging to *Bacilli* in relatively high abundance, although not to the same magnitude as that found by Piao *et al*. (2014). In their meta-analysis, Kim *et al.* (2011) reported that approximately 1.4% of all bacterial sequences belonged to the class *Bacilli*, while Jose *et al*. (2017) found this class to account for more than 7% of combined bacterial and archaeal reads from metagenomic data collected from the rumen fluid of Holstein-Friesian crossbred steers.

Evidence supporting the enriched *Rummeliibacillus* species' role in fiber degradation is supported from the results of our metagenomic analysis. A variety of glycoside hydrolases with known cellulolytic capabilities have been discovered from members of *Firmicutes*. While three GHs from the enriched *Rummeliibacillus* were assigned to a specific family (GH9), others that were not assigned to a previously designated group included cellulase, β-glucanase, and endoglucanase. GH9 members are commonly found in rumen meta- studies and include known cellulases. Dai *et al*. (2015) found this family of GHs to account for 13.71% of the total GHs, and is most heavily associated with *Firmicutes* relative to *Bacteroidetes* and *Fibrobacteres*, with percentages ranging from 1% to greater than 10% of the total plant cell wall polysaccharide (PCWP) degrading domains encoded by the respective genomes of the given genera. In addition,

Brulc *et al*. (2009) found GH9 to have a mean relative abundance of 1.06% compared to all GHs found in their analysis.

Additional downstream cellulose-degrading enzymes were found in the *Rummeliibacillus*-related OTU, including two from known GH families (GH1 and GH3), as well as a variety of unassigned enzymes including β-glucosidase, cellobiose phosphorylase, and aryl-phospho-beta-D-glucosidase. GH1 and GH3 are known oligosaccharide degrading enzymes (β-glucosidases) and are commonly found in rumen metagenomic/transcriptomic studies. Dai *et al*. (2015) showed that GH1 and GH3 accounted for 5.07% and 12.2% of the total PCWP-targeting GHs in the rumen of Holstein cows, respectively. Dai *et al*. (2015) also showed the distribution of different GH families across various genera within *Firmicutes*, *Bacteroidetes*, and *Fibrobacteres*. GH1 was shown to only be present in *Firmicutes*-related bacteria. GH3 was found to range between 0.1% and 10% of the total PCWP degrading domains encoded by the genome of a given genus. In contrast, Brulc *et al*. (2009) found GH1 and GH3 at mean relative abundances of 2.03% and 29.57% of the total GHs found in the rumen of Angus cattle, respectively.

Compared to the *Rummeliibacillus*-enriched OTU, the *Prevotella*-enriched OTU was not as highly represented in the sample used for metagenomic sequencing, rendering fewer *de novo* assembled contigs. A variety of genes were identified in the *Prevotella*-related organism that were predicted to be involved in polysaccharide degradation. While only one unassigned cellulase was identified for the *Prevotella* OTU, a variety of SusD genes were found in the contigs assigned to this organism. Dai *et al*. (2015) described that only organisms belonging to *Bacteroidetes* were found to encode for SusD, which

encodes an integral protein found in polysaccharide utilization loci (PULs) which are gene clusters essential to the degradation of certain complex carbohydrates, including cellulose (Dai *et al.*, 2012). Other genes predicted to be involved in carbohydrate metabolism were identified, including both assigned (GH18, GH25, and GH76) and unassigned glycoside hydrolases, xylanase, alpha-galactosidase, alpha-L-fucosidase, and alpha-rhamnosidase. Together, our results support a role in downstream polysaccharide utilization for this organism and suggest a potential versatility in carbohydrate metabolism.

Certain ruminal bacteria have the ability to convert mono- and disaccharides into glycogen, which is an α-glucan intracellular storage polysaccharide with α-(1,4) and α-(1,6) linkages (Masson & Oxford, 1951; Thomas, 1960). Based on our metagenomic analysis, the enriched organisms likely possess the genetic and metabolic potential to store carbohydrates in the form of glycogen for utilization once the available substrate is exhausted. Indeed, OTU A16-00027 (*Rummeliibacillus* related) possessed a variety of genes involved in glycogen metabolism, including glycogen phosphorylase, glycogen branching protein, 1, 4-alpha-glucan branching enzyme (GlgB), 4-alpha-glucanotransferase, glycogen synthase, glycogen biosynthesis protein (GlgD), and alpha-glucan phosphorylase. It has been described by Denton *et al*. (2014) that ruminal protozoa can sequester sugar away from bacteria by storing carbohydrates in the form of glycogen at a faster rate than bacteria, giving protozoa a distinct evolutionary advantage in the rumen. A similar concept may hold true when bacteria that possess enzymes to not only breakdown cellulose, but also convert it into a reserve carbohydrate, which may give these organisms a competitive advantage over others. Evidence supporting this

hypothesis is also seen in the *Prevotella*-related OTU where the enzyme glycogen synthase was identified.

Typically, ruminal proteolytic bacteria metabolize dietary protein to ammonia, providing cellulolytic bacteria with a nitrogen source for protein synthesis (McSweeney & Mackie, 2012). However, proteolytic activity has been described in some well-characterized cellulolytic bacteria including *P. ruminicola* and *Butyrivibrio fibrosolvens* (Hazlewood *et al.*, 1983). Genes involved in amino acid acquisition (i.e. proteases and membrane transporters) were identified in the enriched OTUs. A high number of peptidases involved in the hydrolysis of various di- and tripeptides (including cysteine and serine types) were found in both enriched OTUs. These have also been reported as the most common rumen bacterial proteases (McSweeney & Mackie, 2012). Interestingly, three distinct ORFs found in the *Rummeliibacillus*-related organism corresponded to peptidase M23, a β-lytic metallopeptidase family that contains endopeptidases that lyse bacterial cell wall peptidoglycans (Grabowska *et al.*, 2015). With this capability, it is possible that this organism could hydrolyze cell wall components from dead bacterial cells in culture.

Thus, we hypothesize that two main characteristics allowed for these microorganisms to be enriched in our system, in addition to the ability to metabolize cellulose. First, the ability to store glucose as glycogen would allow an energy reserve once cellulose in the culture became exhausted. Second, the ability to metabolize peptidoglycans would allow the organism to use dead cells as a source of amino acids.

Conclusion

The use of bovine rumen fluid cultures supplemented with cellulose allowed the enrichment of two previously uncharacterized bacterial species-level OTUs. Metagenomic analysis revealed a variety of genes predicted to function in cellulose metabolism in both enriched OTUs, including cellulolytic enzymes and cellulose-binding proteins. Further analysis of these organisms' genomes may give insight into the physiochemical conditions required to isolate them in pure culture. In striving towards a better understanding of ruminal cellulose metabolism, investigation of individual populations of cellulolytic bacteria offers an alternative approach, with the potential to offer new insight on strategies to enhance feed efficiency in ruminant livestock.

Literature Cited

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Aziz RK, Bartels D, Best AA*, et al.* (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Bayer EA, Lamed R, White BA & Flint HJ (2008) From cellulosomes to cellulosomics. *Chemical Record (New York, NY)* **8**: 364-377.

Brulc JM, Antonopoulos DA, Miller ME*, et al.* (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences of the USA* **106**: 1948-1953.

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

Cunningham C, McPherson CA, Martin J, Harris WJ & Flint HJ (1991) Sequence of a cellulase gene from the rumen anaerobe Ruminococcus flavefaciens 17. *Molecular & General Genetics* **228**: 320-323.

da Mota FF, Vollu RE, Jurelevicius D & Seldin L (2016) Whole-genome sequence of Rummeliibacillus stabekisii strain PP9 isolated from antarctic soil. *Genome Announcements* **4**.

Dai X, Tian Y, Li J*, et al.* (2015) Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in cow rumen. *Applied and Environmental Microbiology* **81**: 1375-1386.

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

Denton BL, Diese DE, Firkins JL & Hackmann TJ (2015) Accumulation of reserve carbohydrate by protozoa and bacteria in competition for glucose. *Applied and Environmental Microbiology*.

Doerner KC & White BA (1990) Assessment of the endo-1,4-beta-glucanase components of Ruminococcus flavefaciens FD-1. *Applied and Environmental Microbiology* **56**: 1844-1850.

Edwards U, Rogall T, Bloecker H, Emde M & Boettger EC (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research* **17**: 7843–7854.

Fernando SC, Purvis HT, Najar FZ, Sukharnikov LO, Krehbiel CR, Nagaraja TG, Roe BA & Desilva U (2010) Rumen microbial population dynamics during adaptation to a high-grain diet. *Applied and Environmental Microbiology* **76**: 7482-7490.

Forsberg CW, Cheng K-J & White BA (1997) Polysaccharide Degradation in the Rumen and Large Intestine. *Gastrointestinal Microbiology: Volume 1 Gastrointestinal Ecosystems and Fermentations,* (Mackie RI & White BA, eds.), pp. 319-379. Springer US, Boston, MA.

Gardner RM, Doerner KC & White BA (1987) Purification and characterization of an exo-beta-1,4-glucanase from Ruminococcus flavefaciens FD-1. *Journal of Bacteriology* **169**: 4581-4588.

Grabowska M, Jagielska E, Czapinska H, Bochtler M & Sabala I (2015) High resolution structure of an M23 peptidase with a substrate analogue. *Scientific Reports* **5**: 14833.

Hazlewood G, Orpin C, Greenwood Y & ME B (1983) Isolation of proteolytic rumen bacteria by use of selective medium containing leaf fraction 1 protein (ribulosebisphosphate carboxylase). *Applied and Environmental Microbiology* **45**: 1780-1784.

Her J & Kim J (2013) Rummeliibacillus suwonensis sp. nov., isolated from soil collected in a mountain area of South Korea. *Journal of Microbiology (Seoul, Korea)* **51**: 268-272.

Hess M, Sczyrba A, Egan R*, et al.* (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science (New York, NY)* **331**: 463–467.

Howard GT & White BA (1988) Molecular cloning and expression of cellulase genes from Ruminococcus albus 8 in Escherichia coli bacteriophage lambda. *Applied and Environmental Microbiology* **54**: 1752-1755.

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

Johnson DR, Goldschmidt F, Lilja EE & Ackermann M (2012) Metabolic specialization and the assembly of microbial communities. *ISME Journal* **6**: 1985-1991.

Jose VL, Appoothy T, More RP & Arun AS (2017) Metagenomic insights into the rumen microbial fibrolytic enzymes in Indian crossbred cattle fed finger millet straw. *AMB Express* **7**: 13.

Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**: 27-30.

Kanno M, Katayama T, Tamaki H*, et al.* (2013) Isolation of butanol- and isobutanol-tolerant bacteria and physiological characterization of their butanol tolerance. *Applied and Environmental Microbiology* **79**: 6998-7005.

Kim M, Morrison M & Yu Z (2011) Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiological Ecology* **76**: 49–63.

Koike S, Yoshitani S, Kobayashi Y & Tanaka K (2003) Phylogenetic analysis of fiber-associated rumen bacterial community and PCR detection of uncultured bacteria. *FEMS Microbiology Letters* **229**: 23-30.

Kourtev PS, Nakatsu CH & Konopka A (2006) Responses of the anaerobic bacterial community to addition of organic C in chromium(VI)- and iron(III)-amended microcosms. *Applied and Environmental Microbiology* **72**: 628-637.

Krause DO, Denman SE, Mackie RI, Morrison M, Rae AL, Attwood GT & McSweeney CS (2003) Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. *FEMS Microbiology Reviews* **27**: 663-693.

Lane D, Pace B, Olsen GJ, Stahl DA, Sogin ML & Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the USA* **82**: 6955-6959.

Ma X, Hu X, Liu L, Li X, Ma Z, Chen J & Wei X (2017) The quality changes and microflora analysis of commercial instant soya milk. *Food Science & Nutrition* **5**: 123-130.

Macfarlane GT & Macfarlane S (1993) Factors affecting fermentation reactions in the large bowel. *The Proceedings of the Nutrition Society* **52**: 367-373.

Masson FM & Oxford AE (1951) The Action of the Ciliates of the Sheep's Rumen upon Various Water-soluble Carbohydrates, Including Polysaccharides. *Journal of General Microbiology* **5**: 664-672.

McSweeney C & Mackie R (2012) Micro-organisms and ruminant digestion: state of knowledge, trends and future prospects. Background study paper No. 61. Food Agrie. Org. United Nations, Rome, Italy.

Miyagi T, Javorsky P, Pristas P, Karita S, Sakka K & Ohmiya K (1998) Partial purification and characterization of RalF40I, a class II restriction endonuclease from Ruminococcus albus F-40, which recognizes and cleaves 5'-/GATC-3'. *FEMS Microbiology Letters* **164**: 215-218.

Naas AE, Mackenzie AK, Mravec J, Schückel J, Willats WG, Eijsink VG & Pope PB (2014) Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio* **5**: e01401.

Nyonyo T, Shinkai T, Tajima A & Mitsumori M (2013) Effect of media composition, including gelling agents, on isolation of previously uncultured rumen bacteria. . *Letters in Applied Microbiology* **56**: 63-70.

Ohara H, Karita S, Kimura T, Sakka K & Ohmiya K (2000) Characterization of the cellulolytic complex (cellulosome) from Ruminococcus albus. *Bioscience, Biotechnology, and Biochemistry* **64**: 254-260.

Piao H, Lachman M, Malfatti S, Sczyrba A, Knierim B, Auer M, Tringe SG, Mackie RI, Yeoman CJ & Hess M (2014) Temporal dynamics of fibrolytic and methanogenic rumen microorganisms during in situ incubation of switchgrass determined by 16S rRNA gene profiling. *Frontiers in Microbiology* **5**: 307.

Rincon MT, McCrae SI, Kirby J, Scott KP & Flint HJ (2001) EndB, a multidomain family 44 cellulase from Ruminococcus flavefaciens 17, binds to cellulose via a novel cellulose-binding module and to another R. flavefaciens protein via a dockerin domain. *Applied and Environmental Microbiology* **67**: 4426-4431.

Russell JB, Muck RE & Weimer PJ (2009) Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. *FEMS Microbiology Ecology* **67**: 183-197.

Schloss PD, Westcott SL, Ryabin T*, et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**: 7537-7541.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ & Birol I (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research* **19**: 1117-1123.

Sirohi SK, Singh N, Dagar SS & Puniya AK (2012) Molecular tools for deciphering the microbial community structure and diversity in rumen ecosystem. *Applied Microbiology and Biotechnology* **95**: 1135-1154.

St-Pierre B & Wright A-DG (2015) Investigation of bacterial and methanogen community composition and diversity in full-scale anaerobic manure digesters. *BAOJ Microbiology* **1**.

Thomas G (1960) Metabolism of the soluble carbohydrates of grasses in the rumen of the sheep. *The Journal of Agricultural Science* **54**: 360-372.

Thornton PK (2010) Livestock production: recent trends, future prospects. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **365**: 2853-2867.

Vaishampayan P, Miyashita M, Ohnishi A, Satomi M, Rooney A, La Duc MT & Venkateswaran K (2009) Description of Rummeliibacillus stabekisii gen. nov., sp. nov. and reclassification of Bacillus pycnus Nakamura et al. 2002 as Rummeliibacillus pycnus comb. nov. *International Journal of Systematic and Evolutionary Microbiology* **59**: 1094-1099.

Van Soest PJ (1994) Nutrional ecology of the ruminant (second edition). 476 pages.

Varel VH & Dehority BA (1989) Ruminal cellulolytic bacteria and protozoa from bison, cattle-bison hybrids, and cattle fed three alfalfa-corn diets. *Applied and Environmental Microbiology* **55**: 148-153.

Vercoe PE, Spight DH & White BA (1995a) Nucleotide sequence and transcriptional analysis of the celD beta-glucanase gene from Ruminococcus flavefaciens FD-1. *Canadian Journal of Microbiology* **41**: 27-34.

Vercoe PE, Finks JL & White BA (1995b) DNA sequence and transcriptional characterization of a beta-glucanase gene (celB) from Ruminococcus flavefaciens FD-1. *Canadian Journal of Microbiology* **41**: 869-876.

Wang L, Hatem A, Catalyurek UV, Morrison M & Yu Z (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PloS One* **8**: e78507.

Wang Q, Garrity GM, Tiedje JM & Cole JR (2007) Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**: 5261-5267.

Wang W, Reid SJ & Thomson JA (1993) Transcriptional regulation of an endoglucanase and a cellodextrinase gene in Ruminococcus flavefaciens FD-1. *Journal of General Microbiology* **139 Pt 6**: 1219-1226.

Wang WY & Thomson JA (1990) Nucleotide sequence of the celA gene encoding a cellodextrinase of Ruminococcus flavefaciens FD-1. *Molecular & General Genetics* **222**: 265-269.

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

Yue ZB, Li WW & Yu HQ (2013) Application of rumen microorganisms for anaerobic bioconversion of lignocellulosic biomass. *Bioresource Technology* **128**: 738-744.

Figure 3.1. Taxonomy-based bacterial community profile of rumen sample prior to cellulose enrichment. Values are expressed as relative abundance (%). Other phyla category include *Proteobacteria*, *Fibrobacteres*, *Lentisphaerae, Spirochaetes, Tenericutes, Fusobacteria, Elusimicrobia, Actinomycetes,* and/or *Planctomycetes*. (uncl.= unclassified)

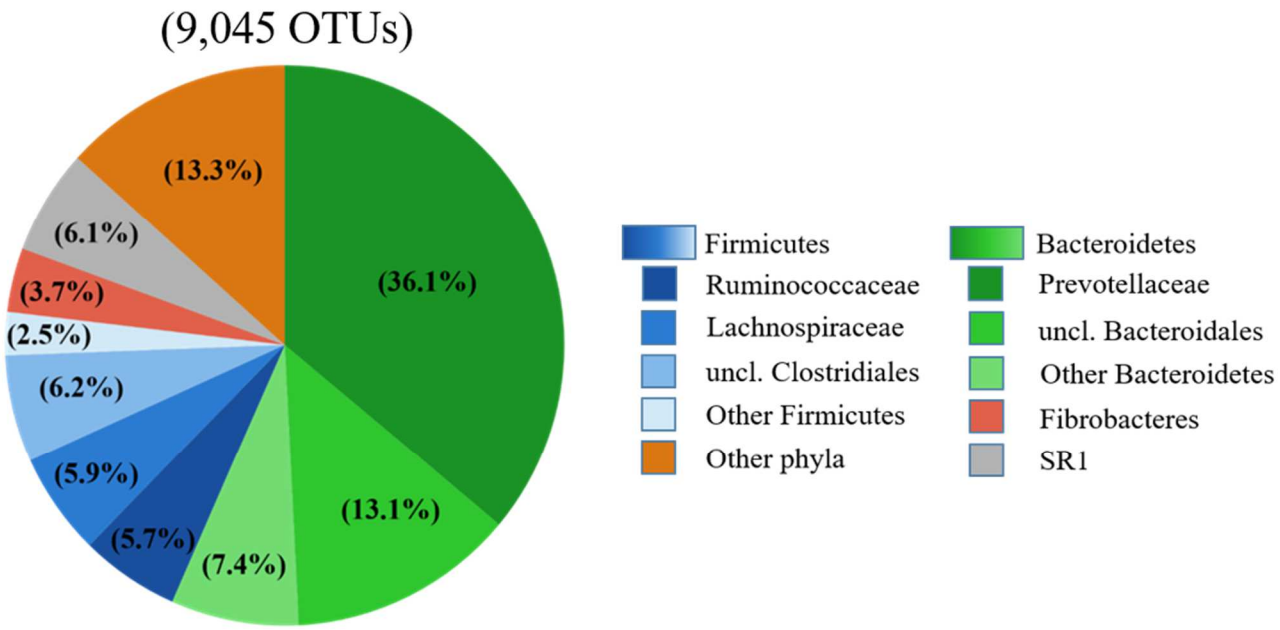

Figure 3.2. Relative abundance of two enriched bacterial species-level OTUs in the presence of cellulose. Relative abundances for the uncultured rumen sample (D0) and cellulose-supplemented cultures (Cel 1, Cel 2, and Cel 3) at D7 and D14 are presented as original values, while means of duplicates are shown for control cultures. Predicted taxonomic affiliations for each of the represented OTUs are shown in the figure legend as determined by RDP Classifier and BLAST.

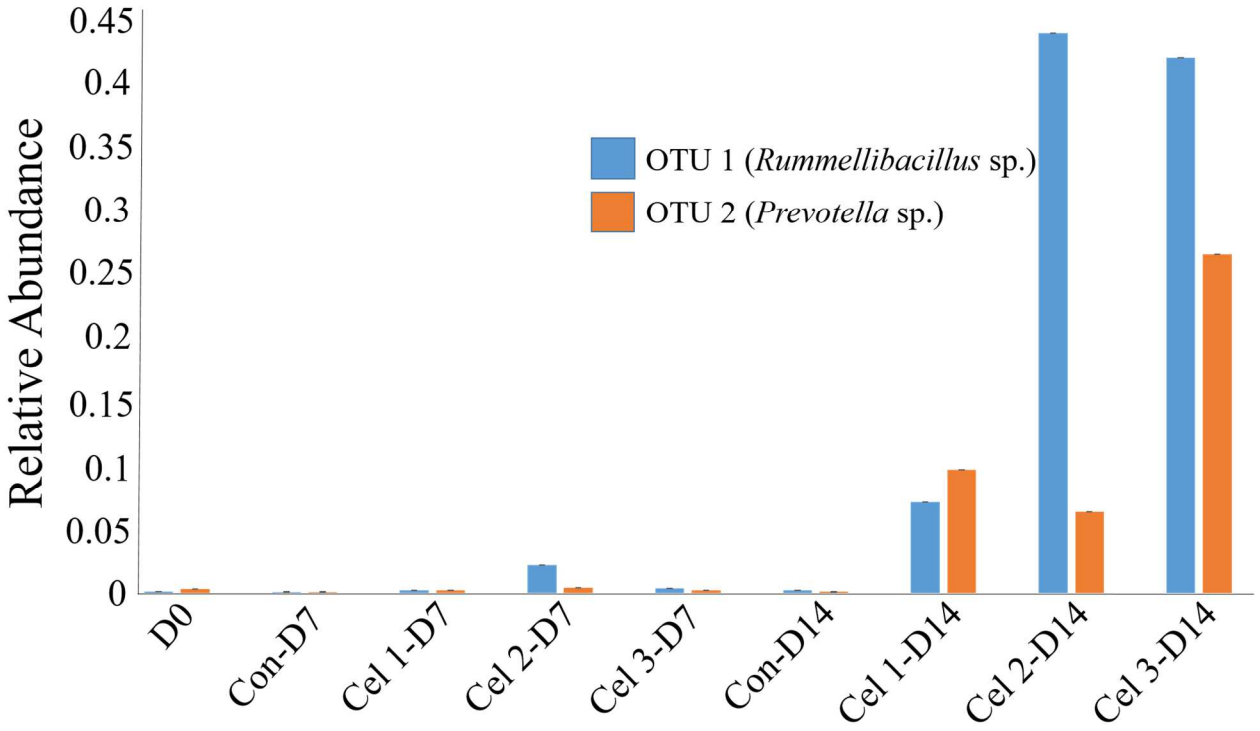

Figure 3.3. Adapted pathway from KEGG annotation with detected genes involved in cellulose and glycogen metabolism found in *Firmicutes*-enriched organism. 1 = endoglucanase/cellulase; 2 = endoglucanase; 3 = PTS system/cellobiose-specific IIA component; 4 = 6-phospho-β-glucosidase; 5 = glucokinase; 6 = phosphoglucomutase; 7 = glucose-1-phosphate adenylyltransferase; 8 = starch synthase; 9 = 1,4-α-glucan-branching-enzyme; 10 = α-amylase; 11 = glycogen phosphorylase; 12 = 4-α-glucanotransferase/maltase-glucanoamylase.

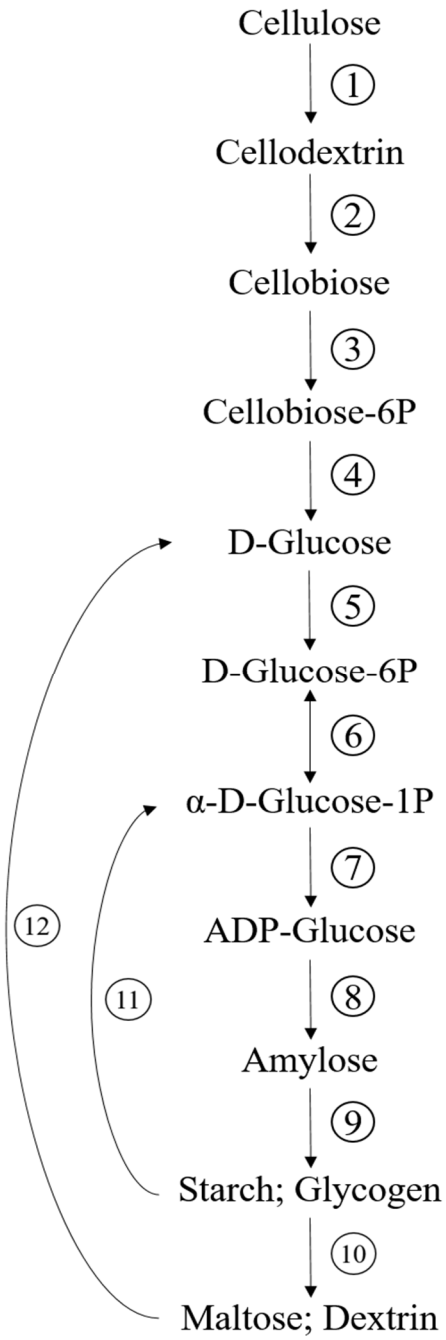

Table 3.1. Relative abundance and RDP Classifier derived taxonomic assignment for highly represented (>1%) OTUs in rumen fluid.

OTU	D0[a]	Taxonomy
A1-01096	4.5	SR1_genera_incertae_sedis
A1-09431	2.6	unclassified_Planctomycetaceae
A1-01719	2.4	unclassified_Bacteria
A1-03801	2.1	unclassified_Bacteroidales
A1-02032	1.8	unclassified_Planctomycetaceae
A1-06496	1.4	Fibrobacter
A1-11322	1.1	SR1_genera_incertae_sedis

[a] Value obtained from single sample

Table 3.2. RAST annotation analysis showing the subsystem distribution for the enriched *Rummeliibacillus* bacteria.

Subsystem	%
None	61.9
Protein Metabolism	6.3
Carbohydrates	4.3
Amino Acids and Derivatives	3.5
DNA Metabolism	2.9
Nucleosides and Nucleotides	2.6
RNA Metabolism	2.6
Membrane Transport	2.1
Cell Division and Cell Cycle	1.7
Stress Response	1.5
Cell Wall and Capsule	1.4
Cofactors, Vitamins, Prosthetic Groups, Pigments	1.3
Virulence, Disease and Defense	1.3
Respiration	1.2
Phages, Prophages, Transposable elements, Plasmids	1.1
Fatty Acids, Lipids, and Isoprenoids	0.8
Motility and Chemotaxis	0.8
Phosphorus Metabolism	0.7
Nitrogen Metabolism	0.5
Regulation and Cell signaling	0.5
Dormancy and Sporulation	0.2
Iron acquisition and metabolism	0.2
Potassium metabolism	0.2
Resistance to antibiotics and toxic compounds	0.2
Miscellaneous	0.1
Sulfur metabolism	0.1

Table 3.3. RAST annotation analysis showing the subsystem distribution for the enriched *Prevotella* bacteria.

Subsytem Category	%
None	64.6
Protein Metabolism	6.6
Carbohydrates	3.9
Amino Acids and Derivatives	2.9
Membrane Transport	2.7
Respiration	2.6
DNA Metabolism	2.4
Nucleosides and Nucleotides	2.3
RNA Metabolism	1.7
Stress Response	1.6
Cofactors, Vitamins, Prosthetic Groups, Pigments	1.5
Fatty Acids, Lipids, and Isoprenoids	1.2
Virulence, Disease and Defense	1.2
Cell Division and Cell Cycle	1
Cell Wall and Capsule	1
Iron acquisition and metabolism	0.5
Nitrogen Metabolism	0.5
Phages, Prophages, Transposable elements, Plasmids	0.5
Phosphorus Metabolism	0.4
Regulation and Cell signaling	0.2
Sulfur metabolism	0.2
Dormancy and Sporulation	0.1
Miscellaneous	0.1
Motility and Chemotaxis	0.1
Potassium metabolism	0.1
Resistance to antibiotics and toxic compounds	0.1

Table 3.4. Carbohydrate-utilizing genes from contigs assigned to the enriched *Rummeliibacillus* bacteria.

Name	Length (aa)	% ID (% Homology)	Gene
Contig_81-ORF84	321	37 (53)	carbohydrate-binding family 9
Contig_127-ORF46	360	74 (86)	cellulase M and related proteins
Contig_382-ORF73	538	71 (83)	6-phospho-beta-glucosidase/cellulase
Contig_382-ORF74	538	70 (84)	beta-glucosidase
Contig_443-ORF40	760	39 (56)	glycosyl hydrolase family 9, cellulase
Contig_443-ORF41	760	39 (55)	beta-glucanase/Beta-glucan synthetase
Contig_443-ORF42	760	37 (53)	beta-glucanase/Beta-glucan synthetase
Contig_443-ORF43	760	38 (57)	glycoside hydrolase family 9
Contig_443-ORF44	760	36 (53)	glycoside hydrolase family 9
Contig_193-ORF1	488	76 (85)	glycosyl hydrolase family 1
Contig_366-ORF48	391	61 (76)	glycosyl hydrolase/glycosyl hydrolase family 3 protein
Contig_254-ORF86	373	67 (82)	endoglucanase
Contig_301-ORF155	508	53 (71)	alpha-L-fucosidase
Contig_57-ORF18	810	66 (83)	glycogen phosphorylase
Contig_92-ORF62	752	70 (83)	phosphorylase, glycogen/starch/alpha-glucan family
Contig_123-ORF41	747	65 (76)	glycogen/starch/alpha-glucan phosphorylase
Contig_371-ORF31	660	74 (84)	glycogen branching protein
Contig_57-ORF75	633	64 (78)	glycogen branching protein
Contig_75-ORF128	575	69 (80)	1 4-alpha-glucan branching enzyme GlgB
Contig_325-ORF34	505	58 (78)	4-alpha-glucanotransferase
Contig_57-ORF62	484	88 (94)	glycogen synthase
Contig_75-ORF131	426	61 (77)	glycogen/starch synthase, ADP-glucose type
Contig_75-ORF78	370	65 (77)	glycogen biosynthesis protein GlgD
Contig_46-ORF106	464	63 (75)	glycoside hydrolase
Contig_308-ORF169	574	57 (73)	glycosyl hydrolase family 2, sugar binding domain protein
Contig_388-ORF292	550	65 (74)	putative glycoside hydrolase
Contig_388-ORF293	550	64 (76)	glycoside hydrolase
Contig_366-ORF221	627	60 (73)	putative glycoside hydrolase
Contig_366-ORF224	627	58 (72)	putative glycoside hydrolase
Contig_102-ORF155	516	56 (64)	Putative glycoside hydrolase
Contig_356-ORF50	381	78 (90)	glycosyl transferase
Contig_195-ORF76	389	64 (78)	glycosyl transferase family 1/1,2-diacylglycerol 3-glucosyltransferase
Contig_167-ORF88	525	52 (68)	glycosyl hydrolase family 25
Contig_127-ORF4	837	78 (86)	Cellobiose phosphorylase

Contig_193-ORF2	488	76 (84)	aryl-phospho-beta-D-glucosidase
Contig_193-ORF3	488	76 (85)	aryl-phospho-beta-D-glucosidase
Contig_186-ORF30	465	65 (82)	aryl-phospho-beta-D-glucosidase BglH
Contig_46-ORF107	464	63 (75)	LysM domain protein
Contig_421-ORF208	380	41 (59)	polysaccharide deacetylase
Contig_366-ORF222	627	60 (73)	alpha-amylase
Contig_366-ORF223	627	62 (74)	alpha amylase catalytic domain protein
Contig_219-ORF41	176	44 (60)	glycosidases/alpha-amylase
Contig_312-ORF87	454	84 (92)	alpha-glucosidase/alpha-galactosidase
Contig_81-ORF83	321	36 (52)	glycoside hydrolase family 10

Table 3.5. Carbohydrate-utilizing genes from contigs assigned to the enriched *Prevotella* bacteria.

Name	Length (aa)	% ID (% Homology)	Gene
Contig_15-ORF6	631	73 (82)	starch-binding protein/SusD family protein
Contig_490-ORF13	522	62 (78)	starch-binding protein, SusD-like family
Contig_148-ORF15	584	62 (76)	SusD family protein
Contig_218-ORF10	687	62 (78)	RagB/SusD family protein
Contig_147-ORF18	666	66 (82)	starch-binding protein, SusD-like family
Contig_140-ORF28	662	94 (96)	glycan metabolism protein RagB/SusD family protein
Contig_97-ORF29	650	55 (70)	RagB/SusD domain protein
Contig_496-ORF35	624	58 (72)	glycan metabolism protein RagB/RagB/SusD domain
Contig_88-ORF38	507	66 (78)	SusD family
Contig_345-ORF10	549	72 (84)	SusD family protein
Contig_276-ORF24	552	64 (78)	SusD homolog
Contig_358-ORF1	600	70 (82)	SusD family protein
Contig_98-ORF12	735	73 (83)	glycoside hydrolase/cellulase
Contig_98-ORF13	735	73 (83)	alpha-L-fucosidase
Contig_98-ORF14	735	63 (76)	alpha-L-fucosidase
Contig_480-ORF28	272	57 (74)	glycogen synthase
Contig_48-ORF1	323	57 (69)	glycoside hydrolase clan GH-D
Contig_345-ORF11	564	64 (79)	glycoside hydrolase
Contig_297-ORF19	488	53 (68)	glycosyl hydrolase
Contig_430-ORF16	584	47 (61)	glycosyl hydrolase family 18
Contig_430-ORF17	584	46 (61)	glycosyl hydrolase family 18
Contig_297-ORF18	488	55 (72)	hydrolase
Contig_432-ORF14	290	56 (71)	glycosyl hydrolase family 25
Contig_297-ORF20	488	57 (71)	glycosyl hydrolase family 76
Contig_35-ORF3	683	53 (68)	alpha-galactosidase (Melibiase)
Contig_35-ORF4	683	52 (66)	alpha-galactosidase
Contig_35-ORF5	683	52 (66)	alpha-galactosidase (Melibiase)
Contig_35-ORF6	683	51 (65)	alpha-galactosidase
Contig_232-ORF18	414	64 (76)	alpha-rhamnosidase
Contig_383-ORF1	451	73 (85)	beta-galactosidase/beta-glucuronidase
Contig_88-ORF28	496	64 (75)	o-Glycosyl hydrolase
Contig_398-ORF11	421	51 (69)	putative glycoside hydrolase xylanase
Contig_284-ORF34	436	58 (73)	glycosyltransferase/glycoside hydrolase, family 10

Table 3.6. Genes involved in the breakdown of protein from the contigs assigned to the enriched *Rummeliibacillus* bacteria.

Name	Length (aa)	% ID (% Homology)	Gene
Contig_241-ORF465	465	74 (86)	aminopeptidase
Contig_351-ORF11	487	67 (84)	aminoacyl-histidine dipeptidase
Contig_443-ORF146	475	38 (57)	aminoacyl-histidine dipeptidase
Contig_497-ORF143	492	52 (69)	aminoacyl-histidine dipeptidase
Contig_7-ORF17	484	45 (60)	aminoacyl-histidine dipeptidase
Contig_30-ORF97	346	53 (70)	aminopeptidase
Contig_251-ORF21	460	60 (76)	aminopeptidase 1
Contig_466-ORF27	440	46 (66)	aminopeptidase C
Contig_466-ORF28	450	57 (75)	aminopeptidase C
Contig_217-ORF19	435	66 (81)	aminopeptidase I zinc metalloprotease (M18)
Contig_355-ORF51	463	71 (82)	aminopeptidase I zinc metalloprotease (M18)
Contig_469-ORF13	440	60 (74)	aminopeptidase P domain protein
Contig_39-ORF369	327	91 (96)	beta-aspartyl peptidase
Contig_366-ORF434	236	63 (79)	CAAX amino protease
Contig_251-ORF74	456	54 (72)	dipeptidase
Contig_92-ORF83	454	69 (82)	dipeptidase
Contig_449-ORF230	484	61 (78)	dipeptidase B
Contig_366-ORF245	344	78 (88)	endopeptidase
Contig_285-ORF74	783	63 (76)	endopeptidase O
Contig_369-ORF14	1507	70 (83)	glutamate synthase
Contig_96-ORF35	612	76 (86)	glutamate synthase
Contig_191-ORF68	802	75 (87)	lon protease
Contig_135-ORF14	437	68 (79)	m18 family aminopeptidase
Contig_40-ORF37	642	100 (100)	M23/M37 family Peptidase
Contig_39-ORF310	358	69 (83)	m42 glutamyl aminopeptidase
Contig_158-ORF55	600	59 (72)	metalloendopeptidase
Contig_110-ORF43	296	100 (100)	methionine aminopeptidase
Contig_263-ORF137	601	63 (79)	oligoendopeptidase F
Contig_461-ORF35	597	71 (84)	oligoendopeptidase F
Contig_105-ORF136	479	40 (60)	peptidase
Contig_451-ORF44	511	51 (69)	peptidase
Contig_443-ORF66	438	63 (77)	peptidase C1
Contig_481-ORF32	459	46 (63)	peptidase C1
Contig_443-ORF46	783	43 (60)	peptidase C11
Contig_259-ORF17	338	54 (68)	peptidase C60B family protein

Contig_100-ORF31	491	65 (78)	peptidase C69
Contig_326-ORF4	423	74 (86)	peptidase M16
Contig_326-ORF8	417	68 (83)	peptidase M16
Contig_331-ORF49	956	49 (66)	peptidase M16 inactive domain protein
Contig_228-ORF33	387	53 (73)	peptidase M20
Contig_362-ORF31	458	89 (95)	peptidase M20
Contig_228-ORF17	393	49 (68)	peptidase M20/M25/M40 family protein
Contig_241-ORF524	420	51 (71)	peptidase M23
Contig_3-ORF72	502	99 (99)	peptidase M23
Contig_498-ORF29	354	76 (86)	peptidase M23
Contig_46-ORF43	355	51 (69)	peptidase M24
Contig_443-ORF68	413	50 (68)	peptidase M29
Contig_96-ORF135	412	71 (84)	peptidase M29
Contig_127-ORF46	360	75 (83)	peptidase M42
Contig_191-ORF75	357	65 (79)	peptidase M50
Contig_280-ORF19	213	44 (57)	peptidase M50
Contig_241-ORF468	253	69 (80)	peptidase S14
Contig_195-ORF131	574	60 (77)	peptidase S8 and S53 subtilisin kexin sedolisin
Contig_422-ORF48	662	42 (61)	peptidase S9
Contig_422-ORF40	294	57 (72)	Peptidase T
Contig_499-ORF17	406	66 (77)	peptidase T
Contig_245-ORF119	408	76 (87)	peptidase U32
Contig_366-ORF33	703	69 (79)	peptidase U32
Contig_361-ORF1	542	64 (77)	peptidase U62
Contig_447-ORF11	428	41 (63)	peptidase, M16 family
Contig_451-ORF82	480	44 (64)	peptidase, M23 family
Contig_57-ORF1	515	48 (65)	peptidase, M23 family
Contig_323-ORF18	446	34 (61)	peptidase, M23 family, partial
Contig_50-ORF20	745	53 (70)	peptidase, U32 family
Contig_178-ORF95	360	63 (77)	proline dipeptidase
Contig_178-ORF28	293	66 (81)	proline iminopeptidase
Contig_282-ORF110	406	65 (80)	protease
Contig_379-ORF168	482	74 (85)	serine carboxypeptidase
Contig_259-ORF51	501	61 (73)	serine protease
Contig_421-ORF210	437	53 (67)	serine protease
Contig_199-ORF137	470	40 (58)	serine-type D-Ala-D-Ala carboxypeptidase
Contig_26-ORF18	450	48 (65)	serine-type D-Ala-D-Ala carboxypeptidase
Contig_371-ORF72	417	52 (68)	xaa-pro aminopeptidase
Contig_435-ORF113	1082	49 (65)	Zinc carboxypeptidase
Contig_479-ORF12	962	41 (59)	zinc-dependent peptidase

Contig_210-ORF67	459	46 (63)	Zn-dependent protease

Table 3.7. Genes involved in the breakdown of protein from the contigs assigned to the enriched *Prevotella* bacteria.

Name	Length (aa)	% ID (% Homology)	Gene
Contig_383-ORF40	527	70 (86)	c-terminal processing peptidase-3
Contig_432-ORF24	395	72 (83)	aminopeptidase C
Contig_485-ORF59	1495	78 (88)	glutamate synthase large subunit
Contig_60-ORF33	802	82 (92)	lon protease
Contig_427-ORF31	316	59 (76)	metalloendopeptidase-like membrane protein
Contig_196-ORF36	434	82 (92)	peptidase
Contig_266-ORF23	644	56 (72)	peptidase
Contig_20-ORF5	401	31 (48)	peptidase C10 family protein
Contig_56-ORF11	437	77 (91)	peptidase M16
Contig_244-ORF35	848	68 (83)	peptidase M16 inactive domain protein
Contig_372-ORF33	368	59 (75)	peptidase M28
Contig_196-ORF28	706	73 (85)	peptidase S10
Contig_213-ORF67	1089	71 (85)	peptidase S41
Contig_287-ORF6	544	81 (89)	peptidase S41
Contig_296-ORF41	632	60 (72)	peptidase U32 family protein
Contig_287-ORF16	693	82 (91)	peptidyl-dipeptidase Dcp
Contig_111-ORF58	480	80 (89)	putative aminopeptidase
Contig_398-ORF4	349	72 (83)	rhomboid family intramembrane serine protease

Chapter 4

FUTURE DIRECTIONS AND IMPACT OF RESEARCH

Introduction

Since the original work done by Robert Hungate in the 1960's, the field of ruminant microbiology has been a growing field of interest for microbial ecologists. In particular, cellulose metabolism in ruminants is respected by researchers around the world for its unique nature and practical applications. The ability of rumen microorganisms to convert cellulose into more utilizable energy compounds has obvious implications and continues to grab the attention of researchers in many areas of science. In the broad field of animal science, research focusing on improving cellulose digestibility in ruminant species aims to increase overall animal performance (Fox *et al.*, 1995). While traditional practices have primarily relied on colony isolation to study these organisms, recent breakthroughs in DNA sequencing has led to new approaches in studying complex microbial communities, and has strengthened our understanding of the complexity of the rumen ecosystem. However, even with the advances in technologies used to investigate the rumen microbiome, much remains unknown in regards to ruminal cellulose metabolism. The experiments outlined in this thesis aimed at furthering our understanding of bacterial cellulose metabolism in the rumen.

Experimental Findings and Future Directions

In the first set of experiments discussed in this thesis, five species-level OTUs were enriched in bovine rumen fluid cultures supplemented with cellulose. Interestingly, one of these OTUs corresponded to a known cellulolytic species, *Ruminococcus flavefaciens*. Enriching for an organism proven to degrade cellulose provides justification for the validity of the approach used in this experiment, and supports that the other novel OTUs enriched are also involved in cellulose catabolism. However, it is quite possible

that some, or all of these organisms, are degrading products generated from downstream cellulose metabolism, and/or possess the genetic and metabolic capability to store glucose for utilization once all the available substrate is exhausted.

Since function typically cannot be determined through 16S rRNA gene diversity studies, further investigation of the enriched uncharacterized organisms' genome will be needed in order to confirm that these organisms are metabolizing cellulose and what mechanisms they possess for carbohydrate storage. An advantageous direction that could be taken would be to conduct a metagenomic or metatranscriptomic analysis of the enriched samples. On the one hand, a metatranscriptomic analysis would be indicative of metabolically active genes within the microbial community, the difficulty in isolating mRNA can be difficult in prokaryotes, and a "snapshot" of gene expression may not be representative of the overall metabolic activity occurring in rumen fluid cultures. Also, transcriptome reference databases are currently limited in their coverage (Aguiar-Pulido *et al.*, 2016). On the other hand, a metagenomic analysis allows researchers to study the genetic capability of the organisms within an environmental sample. However, analyzing sequenced reads derived from metagenomic samples tends to be more difficult in complex environments, such as the rumen, where the high diversity of organisms poses a challenge in reaching a sequence coverage sufficient for the assembly of contigs long enough to represent the genomes of the organisms present in the sample. For this reason, assembly is easier for organisms that are highly represented in a particular sample. Since the organisms identified in the first study were found in much higher abundance, contig representation for these organisms becomes drastically higher. By generating a higher number of long contigs corresponding to the respective genomes of the enriched OTUs,

we can determine their metabolic potential, i.e. confirm their ability to metabolize specific substrates (e.g. cellulose) and uncover other biochemical activities.

A number of genes associated with polysaccharide metabolism were identified from both organisms, including genes corresponding to glycogen synthesis for storage. Thus, two species-level OTUs found to be enriched in the same culture in the presence of cellulose were further characterized using metagenomics in the second set of experiments discussed in this thesis. While fewer genes involved in glycogen storage were identified for the *Prevotella*-related OTU, all the necessary genes needed to metabolize cellulose into glucose, and glucose into glycogen were found in the *Rummeliibacillus*-related organism. Having the capability to metabolize cellulose and store the resulting freed glucose in the form of glycogen could give this organism a distinct advantage in environments both rich in cellulose and recently exhausted of cellulose or downstream cellulose catabolism byproducts. Also found for both organisms were genes involved in protein degradation and amino acid transport, suggesting that these organisms obtain their amino acids for protein synthesis from exogenous proteins and peptides. While the primary focus of our results from the metagenomic analysis was carbohydrate and protein-related genes, a wealth of other genes were identified in the enriched organisms. Further investigation of all the annotated genes found for each organism may provide further insight on what substrates may be needed in order to isolate these organisms in culture.

Perhaps the most intriguing part about this research was having the opportunity to continually expand on experimental findings. If given more time and money to work on this project, I would continue to supplement rumen fluid cultures with cellulose, hoping

to eventually see select OTUs repeatedly enriched. For each enrichment, I would conduct a metagenomics analysis to determine which genes may be responsible for the organism's ability to thrive under the given nutrient conditions. One aspect I would experiment with would be to dedicate three bioreactors in which cellulose is added to the rumen fluid culture only at the beginning of the trial, similar to what was done in the enrichment experiment outlined in this thesis, and to dedicate three other bioreactors in which cellulose is added at regular intervals throughout the trial duration. I would compare the community 16S rRNA gene profiles between the two different treatment groups and the controls, and determine whether or not a difference in enrichment was observed. If a difference in enrichment was observed, I would conduct metagenomics on both treatment groups to try and identify genetic and metabolic differences between the enriched organisms. Since we would be unable to identify which genes are metabolically active, I would also employ a metatransciptomics analysis for both treatment samples during peak enrichment to determine differences in gene expression. Additionally, I would try transferring a small volume of sample from the treatment cultures after several days of incubation to new bioreactors that contains autoclaved rumen fluid and supplemented cellulose. The goal of this would be to try and achieve increased enrichment. With greater enrichment, we are more likely to assemble larger contigs in a metagenomics analysis, revealing more genetic information for the enriched organism(s).

Identifying novel cellulose metabolizing bacteria was the focal point of this project, however, I am also interested in bacteria able to metabolize downstream by-products of cellulose degradation. Ruminal fluid cultures would initially be supplemented with by-products such as cellodextrin and cellobiose, and enrichment could be compared

to cultures containing the same rumen fluid, except supplemented with cellulose. Differences in enrichment could be used to suggest metabolic capabilities for the enriched OTUs.

Future directions for this experiment could also include cultivation-based work. Using autoclaved rumen fluid from an enriched sample supplemented with cellulose, a synthetic agar-based medium could be synthesized. Under anaerobic conditions, plates would be streaked with the sample containing the enriched organism from the rumen fluid used to make the media. Ideally, colonies of the enriched organism(s) would grow and could be isolated for pure culture DNA extraction.

Impact of Research

The majority of research within the field of ruminant microbiology studies rumen microorganisms as a community, whether it is analyzing microbial diversity using phylogenetic markers, metagenomics searching for carbohydrate-active genes or other broad categories of genes, or metatranscriptomics looking at metabolically active genes within a community (Brulc *et al.*, 2009; Piao *et al.*, 2014; Dai *et al.*, 2015; Huws *et al.*, 2016; Jose *et al.*, 2017). Few studies focus on individual populations of microorganisms, with the exception of rumen bacteria that have already been well-characterized (Wang *et al.*, 2013; Oss *et al.*, 2016). While it is necessary to analyze rumen microbial community responses to specific treatments, more emphasis needs to be made on identifying functionality in individual populations of microbes in order to gain further insight on the organisms responsible for specific substrate degradation. The research outlined in this thesis thus aimed at identifying uncharacterized populations of bacteria involved in

cellulose metabolism, and served as a stride forward towards further characterizing the rumen microbiome.

The onset of next-generation sequencing has revealed the very high diversity of the rumen microbiome, with only a fraction having been cultured as isolates. Proposed explanations regarding the difficulty in culturing most anaerobic organisms include the lack of understanding of optimal physiochemical and environmental conditions for individual organisms, and symbiotic relationships essential to survival (Puspita *et al.*, 2012). A number of cellulolytic bacteria have been identified using cultivation-based methods, including *Fibrobacter succinogenes*, *R. flavefaciens*, *R. albus*, and *P. ruminicola*. However, metagenomics studies have revealed an abundance of cellulose-degrading genes that do not belong to the genome of these previously identified cellulolytic organisms, indicating that a large number of organisms involved in its metabolism remain unidentified (Hess *et al.*, 2011; Wang *et al.*, 2013; Jose *et al.*, 2017). Due to our lack of knowledge on how to cultivate the majority of rumen microorganisms, alternative approaches should be used to single out select populations for an in-depth analysis of complex communities such as the rumen ecosystem.

As was shown in this thesis, nutritional selection can be successfully used to enrich for bacteria based on their genetic and metabolic potential. While its use does not allow isolation of pure bacterial species, enriching for few select species allows us to effectively use metagenomics to analyze the partial genomes of uncharacterized bacteria. As mentioned earlier in this chapter, identifying the genetic potential of these organisms may give greater insight into what may be needed to isolate these novel species. The concept used in this experiment contributes an alternative approach towards assigning

functionality to uncharacterized species of bacteria in the rumen, and is not limited to carbohydrate-type substrates. It can be used for a variety of other substrates that may be fed to ruminants, and could give insight into other functionally different bacterial populations. This research, as it stands, may have only introduced a few candidate cellulose-utilizing bacteria amongst a great number that remain to be identified, but a continuation of this research could lead to the identification of previously uncharacterized mechanisms involved in cellulose metabolism, thus leading to novel strategies that enhance the degradation of cellulose in the rumen, and ultimately resulting in the development of higher feed efficiency strategies.

Literature Cited

Aguiar-Pulido V, Huang W, Suarez-Ulloa V, Cickovski T, Mathee K & Narasimhan G (2016) Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. *Evolutionary Bioinformatics Online* **12**: 5-16.

Brulc JM, Antonopoulos DA, Miller ME*, et al.* (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences of the USA* **106**: 1948-1953.

Dai X, Tian Y, Li J*, et al.* (2015) Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in cow rumen. *Applied and Environmental Microbiology* **81**: 1375-1386.

Fox DG, Barry MC, Pitt RE, Roseler DK & Stone WC (1995) Application of the Cornell Net Carbohydrate and Protein model for cattle consuming forages. *Journal of Animal Science* **73**: 267-277.

Hess M, Sczyrba A, Egan R*, et al.* (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science (New York, NY)* **331**: 463–467.

Huws SA, Edwards JE, Creevey CJ, Rees Stevens P, Lin W, Girdwood SE, Pachebat JA & Kingston-Smith AH (2016) Temporal dynamics of the metabolically active rumen bacteria colonizing fresh perennial ryegrass. *FEMS Microbiology Ecology* **92**: 1.

Jose VL, Appoothy T, More RP & Arun AS (2017) Metagenomic insights into the rumen microbial fibrolytic enzymes in Indian crossbred cattle fed finger millet straw. *AMB Express* **7**: 13.

Oss DB, Ribeiro GO, Jr., Marcondes MI, Yang W, Beauchemin KA, Forster RJ & McAllister TA (2016) Synergism of Cattle and Bison Inoculum on Ruminal Fermentation and Select Bacterial Communities in an Artificial Rumen (Rusitec) Fed a Barley Straw Based Diet. *Frontiers in Microbiology* **7**: 2032.

Piao H, Lachman M, Malfatti S, Sczyrba A, Knierim B, Auer M, Tringe SG, Mackie RI, Yeoman CJ & Hess M (2014) Temporal dynamics of fibrolytic and methanogenic rumen microorganisms during in situ incubation of switchgrass determined by 16S rRNA gene profiling. *Frontiers in Microbiology* **5**: 307.

Puspita ID, Kamagata Y, Tanaka M, Asano K & Nakatsu CH (2012) Are uncultivated bacteria really uncultivable? *Microbes and Environments* **27**: 356-366.

Wang L, Hatem A, Catalyurek UV, Morrison M & Yu Z (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PloS One* **8**: e78507.