Ecology of *Fibrobacter* in the herbivore gut

By

Anthony P. Neumann

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Microbiology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 04/24/2018

The dissertation is approved by the following members of the Final Oral Committee:

Garret Suen, Associate Professor, Bacteriology
Christina Kendziorski, Professor, Biostatistics & Medical Informatics
Katherine D. McMahon, Professor, Civil & Environmental Engineering, Bacteriology
Michael G. Thomas, Professor, Bacteriology
Paul J. Weimer, Associate Professor, Bacteriology

Abstract

Microorganisms responsible for the degradation of plant cell wall polysaccharides in the herbivore gut are of critical importance to the well-being of their hosts because these animals could not efficiently digest their fiber-rich diets without them. Cellulose is typically the most abundant structural polysaccharide in the plant cell wall, and is extremely resistant to enzymatic depolymerization. However, despite this recalcitrance, microorganisms have evolved mechanisms to utilize cellulose as a source of carbon and energy for growth. Important cellulose-degrading specialists in the herbivore gut include members of the bacterial genus *Fibrobacter*. *Fibrobacter* was first formally recognized for its role in the bovine rumen, and has since been implicated in fiber-degradation in the guts of diverse herbivores. Still, much of the phylogenetic diversity among *Fibrobacter* spp. is poorly characterized owing in large part to a lack of representative cultures for study. In this thesis, I have attempted to improve our understanding of the abundance and distribution of these bacteria in their herbivorous hosts using a combination of culture-based and genomic analyses. In chapter two, I describe a novel isolation method used to substantially expand the phylogenetic diversity of axenic *Fibrobacter* cultures available for investigation, and confirm that these cultures are representative of the natural ecology of these bacteria in the herbivore gut through complementary 16S rRNA gene amplicon sequencing of the samples from which they were recovered. In chapter three, I report the results of a comparative analysis of whole-genome sequences generated from 40 *Fibrobacter* strains, which identified enrichments in genes ecoding plant cell wall degrading enzymes in strains of *F. succinogenes* relative to strains of *F. intestinalis*. In chapter four, I describe changes in gene expression observed during growth at different rates on cellulose relative to growth on soluble sugars that further refine our system-level understanding of the unique mechanism of

cellulose degradation and utilization by the *F. succinogenes* type strain, S85. Collectively, this work represents a substantial improvement in our knowledge of these poorly understood bacteria, and provides a rationale for the distinct ecological patterns observed among particular *Fibrobacter* phylotypes in the gastrointestinal tracts of diverse herbivorous animals.

Acknowledgements

This thesis is a result of everyone who has contributed to my growth as a person and as a scientist, which has brought me to where I am today. First and foremost, I thank my parents Steve and Mary Neumann for their unconditional love and encouragement, and for teaching me the value of hard work and education. Thank you to my brother Todd, who I am fortunate to include among my closest of friends. Thanks also to my grandparents, aunts, uncles, and cousins. I am so blessed to have such a close and supportive family. Thank you to my best friends Sam Sebastian and Scott Krumsee for always making me laugh and bringing joy to my life.

Thank you to my first mentors in science Dr. Tom Rehberger and Dr. Greg Siragusa, as well as all of my former coworkers at Agtech, for inspiring me to dedicate my life to science and microbiology in particular. Thank you to my thesis advisor, Dr. Garret Suen, for taking a chance on an unknown Master's student and for providing me with every opportunity to take this work where I wanted to see it go. I will greatly miss the culture of the Suen lab, which is such a safe and nuturing environment to pursue science in. Thank you also to all of the members of the Suen lab whom I've had the honor of working with and learning from over the past years. I have particularly appreciated the opportunity to mentor several talented undergraduate students, including Caroline McCormick, Gaby Moen, and Sofia Magalhaes Moreira. I also wish to express my gratitude to the members of my thesis committee for their valuable feedback, and Dr. Paul Weimer in particular for always being willing to take extra time for me, and for teaching me so much about anaerobic fiber-degrading bacteria. I would like to acknowelge Dr. Bob Kerby as well for his helpful discussions regarding the cultivation of fastidious anaerobes. Thanks to everyone in the Microbiology Doctorol Training Program, especially Cathy Davis Gray for

making this program such a great experience. Thank you to all of the collaborators who participated in this work by collecting samples for the analyses described in chapter two: Dr. Erin Flynn and colleagues at the Henry Vilas Zoo; Beth Rich, Trish Khan, and Tim Wild at the Milwaukee County Zoo; Dr. Jon Roll in the Department of Bacteriology; and Dr. Benjamin Darien at the UW-Madison School of Veterinary Medicine

An extra special acknowledgement is reserved for the love of my life and wife, Dr. Gina Neumann. You are the best partner I could hope to have, and I appreciate every second we have shared as we've pursued this journey together. I am so excited to be starting a new phase of our life together, and can't wait to see where science takes the two of us in the future. Thank you also to Gina's family for supporting us and expressing such an interest in what we do. Finally, thanks to my buddy boy, Dixon, for keeping me company during lonely days of writing.

This thesis is dedicated to the memory of Steve Neumann, Pat Burns, Mike Burns, Matt Burns, and Marty Burns.

Table of Contents

List of Figures

Page

List of Tables

Page

Chapter 1: Introduction

Herbivores share a symbiotic relationship with microorganisms living in their gastrointestinal tract that they depend on to digest the plant cell wall polysaccharides (PCWP) present in their diet (1). Cellulose is typically the most abundant structural polysaccharide in plants, and is resistant to hydrolysis due to its crystalline higher-order structure (2). Nevertheless, certain microorganisms are able to effectively deconstruct the cellulose present in plant matter, and it is believed that these microbes are of fundamental importance to fiber-degradation in environments such as the herbivore gut (3, 4). Ruminant livestock, which have evolved an expanded and multi-chambered foregut to allow for more efficient digestion of their plant-based diet, represent the most well studied model of herbivore host-microbe interactions (5). The rumen is the foregut compartment where the bulk of the hydrolysis and fermentation of plant material occurs, and it is well established that the most abundant cellulose-degrading microorganisms present are bacterial species of *Ruminococcus* and *Fibrobacter* (6).

Representatives of the genus *Fibrobacter* were first conclusively described in the context of the bovine rumen in the mid-twentieth century by Robert Hungate (7). Using novel methods for the isolation of strictly anaerobic cellulose-degrading bacteria, Dr. Hungate purified, from the rumen of cattle, cultures of Gram-negative, non-sporeforming, non-motile, obligately anaerobic rod-shaped bacteria that actively hydrolyzed and fermented cellulose to primarily succinic and acetic acids (8). Shortly thereafter, Bryant provided the original description of what would become the *Fibrobacter* genus type strain S85, which he isolated from the 10^{-8} dilution of rumen contents sampled from a heifer fed alfalfa silage (9). Although these cultures were initially classified to the genus *Bacteroides,* subsequent work has established their proper placement in the genus *Fibrobacter*, a major lineage within an entirely different bacterial phylum, the Fibrobacteres (10). More recently, application of culture-independent methods has produced

molecular evidence for *Fibrobacter* and related members of the Fibrobacteres in the gastrointestinal tracts of diverse mammalian herbivores and termites as well as for free-living populations in landfills and lake sediments, suggesting a role for these bacteria in diverse anaerobic and microaerophilic environments where the decomposition of PCWP occurs (11–16).

1.1 *Fibrobacter* phylogeny

The original ruminal *Fibrobacter* isolates were classified as *Bacteroides succinogenes* because of their small rod-shaped cells, obligately anaerobic nature, failure to produce spores, Gram-negative cell envelope, intestinal habitat, and production of succinate as the major fermentation product (8). However, oligonucleotide cataloging of 16S rRNA and later 16S rRNA gene sequencing provided justification for reclassification as a novel species, designated *Fibrobacter succinogenes*, that is sufficiently phylogenetically divergent to warrant placement in its own bacterial phylum, the Fibrobacteres (10, 17). More recently, the publication of a complete genome for strain S85 as well as a comparative analysis of Fibrobacteres population genomes has reinforced the distinct phylogeny of these bacteria (18, 19).

Currently, there are only two officially described species in the genus *Fibrobacter*: *F. succinogenes* and *F. intestinalis* (10). The *F. intestinalis* type strain NR9 was isolated from the cecum of a rat, and despite considerable divergence in 16S rRNA gene sequence no phenotypic differences have been identified that reliably differentiate the two species (20, 21). Additionally, there are now two formally described subspecies recognized within *F. succinogenes*: *F. succinogenes* subsp. *succinogenes* and *F. succinogenes* subsp. *elongatus*, represented by the rumen strains S85 and HM2, respectively (10). The most striking characteristic differentiating the two subspecies is cell morphology, with cells of *F. succinogenes* subsp. *succinogenes* being

coccoid or lemon-shaped whereas cells of *F. succinogenes* subsp. *elongatus* are slender rods (10, 21). 16S rRNA gene sequencing of other *Fibrobacter* isolates suggests that there are at least two other well resolved phylogenetic groups within the *F. succinogenes* clade and likely several within the *F. intestinalis* clade as well, but a lack of clear phenotypic differences or evidence for distinct ecological niches precludes their reclassification as distinct species or subspecies (Figure 1.1) (21).

Analysis of DNA sequences recovered using culture-independent methods suggests substantially more diversity than is currently represented by isolated cultures (22, 23). This includes evidence for phylogenetically distinct populations in the horse hindgut, the termite gut, ostrich cecum, landfills, and lake sediments (13–15, 24, 25). The recovery of Fibrobacteres population genomes from termite gut, anaerobic digester, and sheep rumen metagenomes has provided the most detailed view of the phylum to date, particularly outside of the genus *Fibrobacter* (19). Phylogenetic modeling of concatenated conserved proteins from these population genomes, the *F. succinogenes* S85 genome, and an outgroup consisting of more than 2,000 bacterial genomes from 33 phyla identified robust monophyly of the Fibrobacteres and representatives of TG3, a poorly characterized phylum-level lineage originally identified in the termite gut that also contains *Chitinivibrio alkaliphilus*, a chitin-degrading species recently isolated from hypersaline soda lake sediment (19, 26). The addition of TG3 to the Fibrobacteres introduced at least one additional class, Chitinivibrionia, to the Fibrobacteres. Moreover, a separate monophyletic cluster containing Fibrobacteres population genomes in class Fibrobacteria was found exclusively in termite guts and designated as the novel order Fibromonadales, containing family Fibromonadaceae, with the most complete representative population genome given the candidatus name, *Fibromonas termitidis* (19).

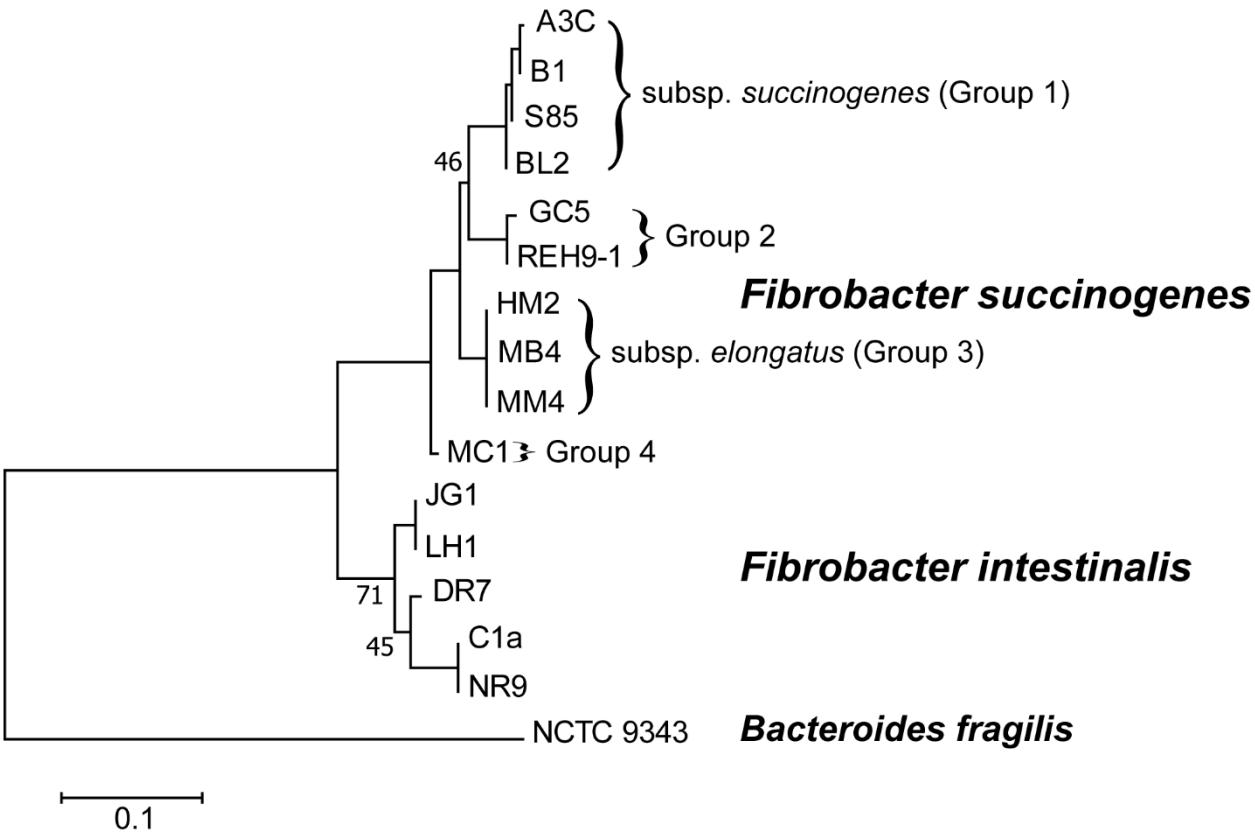


Figure 1.1: Maximum-likelihood phylogeny of *Fibrobacter* spp. The tree was reproduced using 16S rRNA gene sequences originally published by Amann *et al.* (21), with *Bacteroides fragilis* NCTC 9343 (NCBI ref. NR_074784.2) included as the outgroup. The scale of the branch lengths is in units of the number of substitutions per site. Branches with < 80% support (100 bootstrap replicates) are labeled.

1.2 *Fibrobacter* physiology

All known strains of *Fibrobacter* are able to grow on cellulose and cellobiose with the concomitant production of primarily succinic and acetic acids (8, 9, 20, 27–29). Many strains produce traces of formic acid as well, but the production of H_2, lactic acid, or ethanol has not been reliably demonstrated (8, 9). An absence of growth on glucose has been reported for certain strains, however the vast majority grow readily on this substrate (8, 9, 20). Select strains have also been reported to grow on the soluble sugars: maltose, trehalose, and lactose (8, 9). Although most, if not all, strains solubilize hemicellulose from intact plant cell walls, none have

been demonstrated to grow on soluble pentosans or xylose (9, 30, 31). This characteristic is likely attributable to a lack of genes or activities for xylose permease, xylose isomerase, and xylulokinase (18, 32). CO_2 is required for growth and is consumed by these bacteria in the process of succinic acid production (33). NH_3 is utilized for nitrogen and is required by most strains, although growth is often stimulated in the presence of protein hydrolysates (27, 34). A reduced form of sulfur, *e.g.* cysteine or sulfide, is likely preferred and can serve as the sole source of sulfur (34). All strains characterized to date require one medium length straight-chain, *e.g.* valeric acid, and one branched-chain, *e.g.* isobutyric or 2-methylbutyric acid, fatty acid for the synthesis of long-chain fatty acids and fatty aldehydes for building phospholipids (35–37). A loose association has been demonstrated for the specific vitamin requirements of members of certain phylogenetic clades with strains of *F. succinogenes* subsp. *succinogenes* requiring only biotin for growth while strains of *F. succinogenes* subsp. *elongatus* generally require biotin, *p*-aminobenzoic acid (PABA), and sometimes cyanocobalamin. Strains of *F. intestinalis* typically require cyanocobalamin, PABA, and occasionally biotin and/or thiamine (10).

Fibrobacter metabolism has been most extensively characterized using strain S85, but most features appear to be conserved based on currently available genomes and investigations involving other strains (19, 38). Cells actively import both glucose and cellobiose, apparently via separate mechanisms, using energy from the maintenance of a proton-motive force (PMF) or sodium ion gradient (39, 40). There is no evidence supporting the use of a phosphotransferase system for the uptake of these sugars (39, 41). Glucose sugar phosphates can be generated by either the GTP-dependent glucokinase or cellodextrin phosphorylase (38, 42, 43). Glucose 6-phosphate is catabolized via the Embden-Meyerhof-Parnas (EMP) pathway generating nucleoside triphosphates through substrate-level phosphorylation (SLP), reduced pyridine

nucleotides, and the key metabolic intermediate phosphoenolpyruvate (PEP) (44, 45). Many strains, including S85, lack pyruvate kinase (19). Other intermediates of the EMP pathway are used to produce pentose phosphates for nucleotide biosynthesis via enzymes of the non-oxidative branch of the pentose phosphate pathway (PPP) (32, 45). Transketolase, ribulose-5-phosphate 3-epimerase, ribose-5-phosphate isomerase, and phosphoketolase activities have been demonstrated, but the oxidative branch of the PPP as well as the Entner-Doudoroff pathway are not present (32, 45). The PEP generated from the EMP pathway can be used to generate GTP and oxaloacetate with the incorporation of one molecule of CO_2 (44). Oxaloacetate can be used to generate pyruvate and amino acids or it can be reduced to malate in order to regenerate oxidized pyridine nucleotides for the EMP pathway (18, 44, 46). Most acetyl-CoA is likely produced from the oxidation of pyruvate by pyruvate-flavodoxin oxidoreductase, though many strains also exhibit pyruvate formate lyase activity (44, 45). The reduced flavin nucleotides are reoxidized by a membrane-associated fumarate reductase, producing the major fermentation product succinate (44, 47, 48).

In addition to the reductive branch of the tricarboxylic acid (TCA) cycle, S85, and other Fibrobacteria, possess enzymes of the oxidative branch of the TCA to generate α-ketoglutarate (18, 19). As is typical of strictly anaerobic bacteria, current data suggests that all members of the class Fibrobacteria lack the genes for α-ketoglutarate dehydrogenase and succinyl-CoA synthetase (19). Nitrogen assimilation in S85 appears to occur primarily via glutamate dehydrogenase (amination of α-ketoglutarate) and alanine dehydrogenase (amination of pyruvate) (46). During growth on glucose, glutamate dehydrogenase is expressed regardless of excess or limiting concentrations of NH_3 (46). Glutamine synthetase activity was not detected under these conditions, though genes for this enzyme are predicted from the S85 genome (18,

46). The majority of strains appear to be capable of synthesizing most cellular nitrogen-containing compounds from exogenous NH_3, but cells will take up and preferentially utilize amino acids if they are available (34, 49, 50). A curious feature of the Fibrobacteres genomes currently available is the sporadic distribution of core N_2 fixing genes (18, 19). No strains have demonstrated an ability to fix nitrogen, and none of the genomes examined possess a complete set of components required for even the most minimal of functional nitrogenases making their purpose unclear (19).

F. succinogenes S85 obtains ATP from electron transport in addition to SLP (51). Genes for a membrane-bound ATPase are widely conserved among diverse Fibrobacteres and rumen strains possess b-type cytochromes (18, 19, 52). It is estimated that four molar equivalents of ATP are produced for every mole of glucose consumed by S85, and that the amount of energy obtained from SLP and electron transport are approximately equal (51). One potential coupling site for the formation of the transmembrane electrochemical gradient is the NADH-linked fumarate reductase which can receive electrons from the oxidation of NADH by NADH dehydrogenase, a known proton pump, through cytochrome b and the quinone pool (47). It has also been proposed that not all members of the Fibrobacteres are strict anaerobes due to the presence of a complete set of genes for complexes I and II, cytochrome bd complex, as well as ATP synthase in Fibromonadaceae genomes (19). A functional respiratory chain would be beneficial for these organisms in the microaerophilic environment of the termite gut where they are found. S85 is unable to maintain a proton motive force when the pH falls below 5.7 and this appears to be a major factor in the organism's sensitivity to acidic conditions, further demonstrating the importance of the transmembrane electrochemical gradient for energy and transport related functions in these bacteria (53, 54).

Several unusual physiological features have been reported for *Fibrobacter*, particularly in regards to carbohydrate metabolism. Cells grown on soluble sugars accumulate large cytoplasmic deposits of glycogen, but exhibit rapid death after entering stationary phase (29, 55, 56). Glycogen synthesis is apparently unregulated as it is produced during all growth phases as well as both degraded and synthesized simultaneously in resting cells in the presence of soluble sugar (57). Glucose and cellobiose are metabolized concurrently by resting cells, but glucose is used preferentially for energy generation and glycogen synthesis whereas some cellobiose is diverted towards cellodextrin synthesis (38, 58). The accumulation of cellodextrins is positively correlated with the accumulation of glucose 6-phosphate (58). Futile glycogen cycling and cellodextrin synthesis may act as buffers for excess energy and carbon in these bacteria. When NH_3 is added to resting cells with glucose the glycogen storage is reduced by half and the rate of glucose consumption decreased by 34% (46). Cells growing with excess soluble sugar die rapidly after NH_3 is exhausted, but this result is not observed when cellulose is in excess likely because the cellulase system is very sensitive to feedback inhibition and soluble sugars do not accumulate (59, 60). Resting cells of *Fibrobacter* are also capable of synthesizing maltodextrins and maltodextrin 1-phosphates with degrees of polymerization up to approximately 6 – 7 (61, 62). Both phosphorylated and unphosphorylated forms of maltodextrins are excreted into the external medium while the cellodextrins remain associated with the cytoplasm and periplasm (63). S85 has also been shown to produce cellobionic acid during growth on soluble and insoluble substrates as well as an uncharacterized derivative of glucose 1-phosphate (64). All together, these results suggest an organism that is well adapted to carbon-limited conditions, such as those encountered in the rumen (60).

1.3 *Fibrobacter* ecology

Since the initial isolations from the bovine rumen, evidence for the presence of Fibrobacteres populations in diverse environments has accumulated (22). It is now clear that members of the genus *Fibrobacter* are conserved constituents of the rumen microbiota regardless of ruminant species, and also common inhabitants of the lower gastrointestinal tract in hindgut fermenting mammalian herbivores (12, 20, 24, 28, 65). Additionally, populations of distinct Fibrobacteres lineages have been identified in the guts of wood-feeding higher termites as well as free-living populations in anaerobic digesters, landfills, and lake sediments (13–15, 19). In their native environments, these bacteria primarily exist as members of biofilms that form on the surfaces of the plant matter present, which is consistent with their specialization as degraders of insoluble fibrous substrates (66, 67). Their abundance appears to be correlated with the rate of substrate turnover in a given habitat. In the rumen, one of the most actively fibrolytic environments known, *Fibrobacter* populations typically make up between 0.1 - 1% of the total bacterial community ($> 10^7$ cells per gram) while populations in environments where less turnover of fiber occurs, *e.g.* rat cecum or lake sediment, are at least an order of magnitude lower (14, 65, 68–70). Relative abundances of greater than 10% have been reported for Fibrobacteres populations in the termite gut (71).

Certain phylogenetic groups of Fibrobacteres exhibit an association with particular environments. In the rumen, where *F. succinogenes* is dominant, *F. succinogenes* clades I, which includes the type strain S85, and clade II are the most frequently isolated phylotypes (72). There is some evidence for populations of *F. intestinalis* in the rumen, but they are present at much lower abundance (11). In contrast, *F. intestinalis* is dominant over *F. succinogenes* in the ceca of rats and pigs (16, 20, 28). Culture-independent analyses have identified phylogenetically

distinct *Fibrobacter* populations in the hindgut of horses that, despite their gastrointestinal location, are more closely related to *F. succinogenes* based on 16S rRNA gene sequences (24). Specific functional differences within and between the different phylogenetic lineages of *F. succinogenes* and *F. intestinalis* that explain these environmental associations have generally not been elucidated (21). However, metagenomics has informed specific hypotheses regarding functional differences between the termite-associated genus *Fibromonas* and *Fibrobacter* (19). In addition to the potential for respiration, *Fibromonas* population genomes contain genes predicted to encode components of flagellar motility that are absent in *Fibrobacter* genomes (19). While the ability to respire and respond to chemical gradients is beneficial in the termite gut, these features are less advantageous in the more reduced and mixed environment of the mammalian gastrointestinal tract.

Studies investigating the interactions between *Fibrobacter* and other rumen microbes support their fundamental role as primary fiber-degraders in the rumen (Figure 1.2). Although *Fibrobacter* readily solubilize hemicelluloses and pectin from plant cell walls, they do not take up or utilize many of the sugars found in these polysaccharides (30, 31). As a result, organisms that are well adapted for growth on these substrates benefit from the activity of *Fibrobacter*. Numerous reports have quantified this phenomenon by showing enhanced cell wall degradation and hemicellulose utilization when *Fibrobacter* is co-cultured with either rumen *Prevotella* spp. or rumen *Butyrivibrio* spp., bacteria that are proficient hemicellulose utilizers but generally poor solubilizers of structural polysaccharides (31, 73–77). Evidence that *Fibrobacter* can engage in cross-feeding of other microorganisms has also been demonstrated during growth on cellulose. Non-cellulolytic bacteria that have been shown to form stable co-cultures with *Fibrobacter* grown on cellulose include: *Streptococcus bovis*, *Selenomonas ruminantium*, and spirochetes in

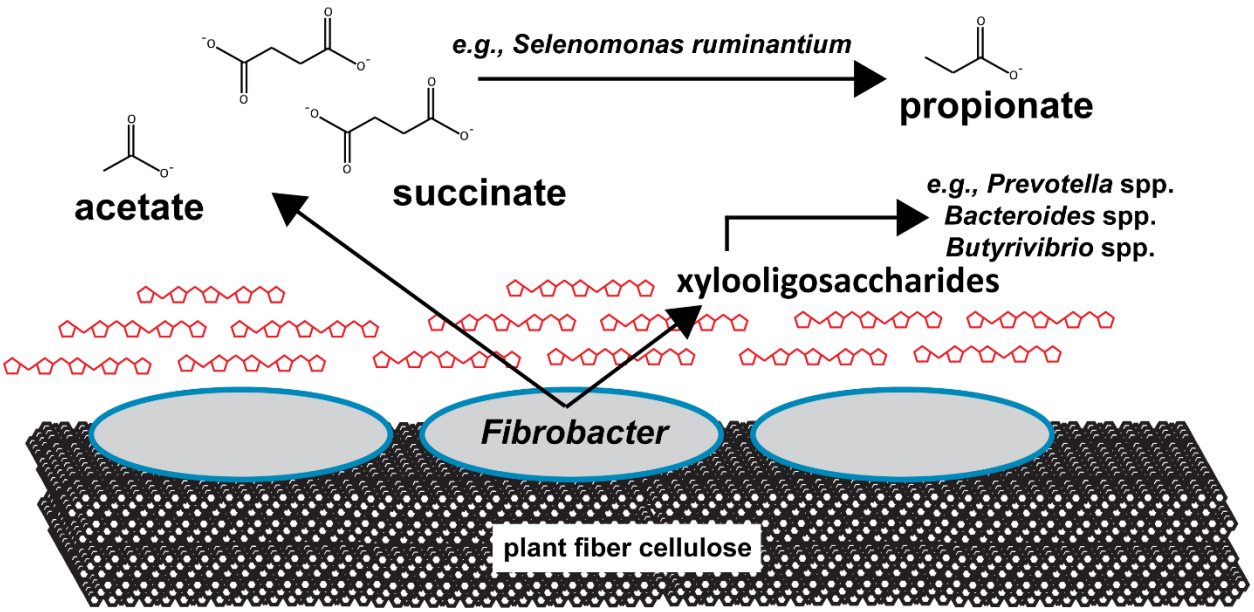


Figure 1.2: *Fibrobacter* cross-feeding in the rumen. *Fibrobacter* does not utilize the xylooligosaccharides it liberates from hemicellulose, leaving them to be catabolized by other members of the rumen microbiota (30, 73). The primary fermentation product produced by *Fibrobacter*, succinate, is decarboxylated by other rumen microbes to produce propionate (80).

the genus *Treponema* (43, 78–81). In these co-cultures, the non-cellulolytics may be scavenging cellodextrins resulting from cellulose hydrolysis, or growing on soluble carbohydrates, *e.g.* glucose and maltose 1-phosphate, observed in small quantities in *Fibrobacter* monocultures during growth on cellulose (64, 82). Succinate, the major fermentation product produced by *Fibrobacter*, is readily decarboxylated to propionate by *S. ruminantium*, providing another potential substrate of cross-feeding (80). This reaction has particular significance in vivo, since succinate does not accumulate in the rumen and propionate is an important substrate of gluconeogenesis for the host (5). *Treponema* are frequent contaminants during attempts to isolate *Fibrobacter*, and it has been hypothesized that by associating with these spirochetes *Fibrobacter* benefit from the spirochete's unique method of motility (81). Other benefits to *Fibrobacter* provided by the microbial community include synthesis of required growth factors

and removal of hydrolysis products which helps limit feedback inhibition of the *Fibrobacter* cellulose-degradation system (35, 74, 83, 84).

Interactions with other cellulose-degrading microorganisms contribute to *Fibrobacter* ecology. *Fibrobacter* and the cellulolytic bacterial species *Ruminococcus flavefaciens* and *Ruminococcus albus* coexist in the rumen, as well as in most other mammalian herbivore gastrointestinal tracts (6, 69). These bacteria are the numerically dominant cellulolytic microbes in these systems, and their individual abundances vary depending on host species, gastrointestinal location, as well as among individual animals, but are usually within an order of magnitude of each other (85). Both groups of bacteria require adherence to insoluble fibrous substrates for efficient growth, and therefore likely engage in fierce competition for colonization *in vivo* (3, 86). Strains of *R. flavefaciens* and *R. albus* have shown greater percentages of cells adhering to cellulose, as well as tighter binding, *in vitro* compared to *F. succinogenes* S85, and, with the exception of glucose, a superior affinity for the products of cellulose hydrolysis (87–89). However, preferences for different colonization sites on more complex plant cell wall substrates have been observed (90, 91). Moreover, while *R. flavefaciens* FD-1 readily outcompetes S85 in cellulose batch cultures, S85 persists in batch cultures grown on alkaline hydroxide-treated wheat straw (92). Furthermore, experiments following the colonization of forage incubated in sacco in the rumen demonstrate greater initial and overall colonization by *Fibrobacter* (93, 94). *F. succinogenes* has been reported to digest more cellulose from intact forages than the cellulolytic ruminococci as well as a greater ability to degrade highly resistant plant cell wall substrates such as undegraded cotton fibers (95, 96). However, variation among *Fibrobacter* strains has also been observed (97). Specific interactions among *Fibrobacter* and cellulolytic rumen fungi are less well understood. *F. succinogenes* S85 can interact positively in the

degradation of ryegrass stem fragments with *Caecomyces* spp., and is negatively affected by certain strains of *Neocallimastix frontalis* (98). Still, other studies have failed to demonstrate any interactions between *Fibrobacter* and anaerobic cellulose-degrading fungi (99–101).

1.4 *Fibrobacter* mechanism of PCWP degradation

Fibrobacter, along with the cellulolytic ruminococci, are the most actively mesophilic cellulolytic bacteria discovered to date (102). Only the thermophile *Clostridium thermocellum* has a higher rate of crystalline cellulose depolymerization (102). The *F. succinogenes* S85 genome contains a high proportion of carbohydrate-active enzymes (CAZymes), and a higher proportion of glycoside hydrolases than even the *C. thermocellum* genome (18, 103). Genes for CAZyme families known to hydrolyze cellulose and xylan are particularly overrepresented compared with other known cellulolytic bacteria (19). The mechanism for hydrolysis appears to be distinct from those used by cellulolytic bacteria in other phyla, but mostly conserved among the Fibrobacteres (18, 19, 104). Members of the Fibrobacteres do not produce cellulosomes, as is typical of most anaerobic cellulose degrading bacteria in the Clostridia (18, 19). Cellulose hydrolysis by *Fibrobacter* is tightly coupled to its energy generating metabolism (8, 84). Growing cells rapidly depolymerize crystalline cellulose, however non-growing cells and cell-free culture filtrates show little hydrolytic activity against this substrate (84, 105). This appears to be at least partially connected to a particular sensitivity to feedback inhibition of the *Fibrobacter* cellulase system (84, 106). Cellobiose and other reducing sugars do not accumulate in batch cultures of *Fibrobacter* grown on excess cellulose after extended periods of incubation suggesting little enzymatic activity after the cessation of growth (8, 27, 64). At least some of the cellulase system appears to be constitutively expressed as no extended lag in growth occurs after

transferring cultures grown on soluble sugar to medium containing cellulose (106). However, there is also evidence that expression of several glycoside hydrolases is enhanced during growth on cellulose versus soluble sugars (107).

Close adherence of *Fibrobacter* cells to insoluble PCWP is a prerequisite for efficient growth and degradation, and substrate surface area appears to be the determinative factor limiting the maximum rate of hydrolysis (102, 108, 109). *F. succinogenes* exhibits similar rates of growth on ball-milled cellulose and cellobiose until all available sites on the cellulose surface are occupied by cells, at which point the growth rate decreases markedly (109). The cell envelope of attached cells conforms closely to the topography of the substrate surface, and digestive pits at the sites of colonization are commonly observed (66, 90). Adherence is disrupted by temperature extremes, and may also be connected to the metabolic activity of the cell (108, 110, 111). Cellobiose, methylcellulose, and carboxymethyl cellulose (CMC) interfere with attachment, and *Fibrobacter* does not grow on cellulose in the presence of methylcellulose, providing additional evidence that attachment is necessary for growth (111, 112). Electron microscopy suggests cells produce a capsule that may be involved in adherence, and cells exhibit increased protein display at the cell surface when grown on cellulose versus soluble sugars (66, 113, 114).

Several proteins that bind cellulose have been identified in the outer membrane (OM) of *F. succinogenes* S85, including proteins that contain the fibro-slime domain (115, 116). The fibro-slime domain is a conserved region of approximately 90 amino acids that is present in large proteins from the slime mold *Dictyostelium discoideum* (115). Ten proteins containing the fibro-slime domain are present in the S85 genome, suggesting multiple duplication events (18). Proteins containing the fibro-slime domain are also present in all Fibrobacteres genomes

analyzed to date, except the chitin degrader *C. alkaliphilus*, but have not been identified in any other bacteria (19). Several fibro-slime domain containing proteins are upregulated in S85 batch cultures grown on cellulose relative to those grown on glucose (117). Glycoside hydrolases that bind cellulose have also been identified in the OM of S85 (115). One of these, Cel51A, also known as EG2 and CelF, has been demonstrated to be responsible for a substantial proportion of the total endoglucanase activity produced by S85 (118). Isolation and characterization of adherence-defective mutants has implicated proteins that may be important for attachment as well (108, 115). A type IV pilin protein that binds cellulose is present in the OM of S85, but absent in mutants that fail to bind to either crystalline or amorphous cellulose (115). Other proteins that bind cellulose and were not observed in mutants contain a tetratricopeptide repeat (TPR) domain (115). The TPR domain has a role in protein-complex formation (119).

Multiple cellulases from CAZyme families GH5, GH9, and GH45 are present in the S85 genome, and many have been purified and characterized (18). Cellulases from GH45 are typical of the Fibrobacteres, but rare for other bacteria (12, 18, 19, 71, 120). Two major endoglucanases account for almost half of the total endoglucanase activity in extracellular culture fluid produced by S85 (118). Endoglucanase 1, also known as Cel9B and CelE, is a GH9 released into the extracellular medium that does not bind Avicel (a microcrystalline cellulose), but has activity against acid-swollen (noncrystalline) cellulose, producing mostly cellobiose (118, 121). Endoglucanase 2, also known as Cel51A and CelF, is the sole GH51 in the S85 genome, is largely cell-associated, binds cellulose, and has high activity against CMC and barley β-glucan, producing cellotetraose (18, 118, 121, 122). Several other secreted enzymes with endoglucanase activity have also been characterized (123). Unlike most cellulolytic microorganisms, the Fibrobacteres appear to lack processive exocellulases since no known strains contain CAZymes

from GH6, GH7, or GH48, families known to perform this function (19). Nevertheless, two periplasmic cellodextrinases, Cel5C and Cel10A, that cleave cellodextrins to cellobiose have been well characterized (124–126). A secreted GH9 enzyme, Cel9D, capable of irreversible hydrolysis of cello-oligosaccharides to glucose appears to also play an important role in cellulose digestion by *F. succinogenes* S85, as do the alternative mechanisms for generating glucose from cellobiose and cellodextrins: cell-associated cellobiase and cellodextrin-phosphorylase in the periplasm (43, 128, 127). It remains to be determined exactly how the cellodextrins generated by extracellular cellulose hydrolysis are imported into the periplasm, but growth on cellulose leads to the upregulation of many predicted transport proteins in the OM (114).

Protuberances evident on the surface of *Fibrobacter* cells suggest that proteins involved in cellulose attachment and hydrolysis form complexes to facilitate degradation (106, 129). *F. succinogenes* glycoside hydrolases exhibit synergy when assayed in combination for the hydrolysis of ball-milled cellulose (130). Many of the proteins present on the cell surface of S85 contain TPR domains, a domain which has been shown to be important for protein-protein interactions (114). TPR domain containing proteins are also found in high numbers in OM vesicles produced by S85, and a multiprotein complex isolated from OM vesicles has demonstrated hydrolytic activity (131). The production of OM vesicles by *F. succinogenes* has been known for some time, but their exact role in plant cell wall degradation is not clear (105, 132). OM vesicles do not appear to be required for *Fibrobacter* growth on cellulose, but are enriched for carbohydrate-related functions relative to all other predicted secreted proteins (131, 133). Hemicellulases seem to be particularly enriched in OM vesicles, and vesicles readily hydrolyze a variety of complex polysaccharide substrates (131). The S85 genome contains many families of known hemicellulases, including members of GH10, GH26, and GH43, and there is

evidence that S85 gives no preference to the degradation of cellulose over hemicellulose when growing on wheat straw, despite relying primarily on the hydrolytic products of the former for growth (18, 134). Several *Fibrobacter* xylanases and carbohydrate esterases, *e.g.* acetylxylan esterase and other hemicellulose debranching enzymes, have been purified and their specific activities confirmed (135–140). The ability to degrade a wide variety of hemicelluloses likely allows *Fibrobacter* to better access the cellulose in plant cell walls which is typically embedded in a chemically diverse matrix of hemicellulose, pectin, and lignin, and OM vesicles may be especially important in this process.

1.5 Motivation and overview

What is the basis of phylogenetic diversity among *Fibrobacter* spp. in the herbivore gut?

Our understanding of the phylogeny of the Fibrobacteres, as well as their abundance and distribution has improved substantially in recent years. Nevertheless, specific explanations for the ecological patterns observed among members of the phylum remain elusive. This is especially the case for the genus *Fibrobacter*, which includes two formally described species, *F. succinogenes* and *F. intestinalis*, that exhibit remarkably similar phenotypes under laboratory conditions given their phylogenetic divergence. Despite these similarities, their environmental distribution is relatively stable and non-random suggesting, that intrinsic features yet to be identified govern selection in their native habitats. A goal of the research described in this thesis was to identify these features with the hope that this knowledge will improve our understanding of diversity among fiber-degrading bacteria of significance for ruminant livestock, as well as among gut microbes generally. A major limitation of previous efforts to explain *Fibrobacter* diversity has been a lack of isolates representative of different phylogenetic lineages and host

species. To address this, I have developed and applied a novel method for recovering *Fibrobacter* isolates. This isolation method in combination with culture-independent 16S rRNA gene amplicon sequencing allowed me to test the hypothesis that *Fibrobacter* spp. in the rumen are distinct from those living in the lower gastrointestinal tract of diverse hindgut fermenting herbivores. To further characterize these differences, the functional potential of the *Fibrobacter* isolates recovered was determined through full genome sequencing. The genomes were then utilized to generate and test specific hypotheses explaining why certain phylotypes are dominant in the rumen while others prevail in the hindgut.

How do *Fibrobacter* spp. degrade fiber?

An additional goal of my thesis research was to generate novel insights into the unique mechanism of PCWP degradation utilized by *Fibrobacter* spp. because a better understanding of this process is necessary in order to fully appreciate fiber degradation in the herbivore gut as well as efficient methods of exploiting plant biomass for industrial applications using microorganisms. Several proteins produced by *F. succinogenes* S85 have demonstrated binding to and/or hydrolysis of PCWP, providing strong evidence for their involvement in the mechanism of fiber degradation. However, the distribution of many these proteins among other members of the genus has not been investigated. I hypothesized that genetic elements important for adherence and hydrolysis would be conserved among diverse *Fibrobacter* strains due to their shared lifestyle as specialized cellulose-degraders, and tested this by mining the genomes of the newly isolated *Fibrobacter* strains for genes predicted to be homologs of binding-proteins and hydrolases previously identified in S85. Another aspect of the cellulase system of S85 that remains poorly characterized is its regulation. In order to address this, RNA-Seq was performed

on cells from continuous cultures grown on either glucose, cellobiose, or cellulose. I hypothesized that genes important for digesting cellulose would be upregulated during growth on this substrate compared to growth on soluble sugars. Novel predictions regarding the mechanism of cellulose degradation were also generated by identifying cellular processes that are co-regulated with previously implicated features of the cellulase system.

1.6 References

1. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol 6:121–131.
2. O'Sullivan AC. 1997. Cellulose: the structure slowly unravels. Cellulose 4:173–207.
3. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577.
4. Leschine SB. 1995. Cellulose degradation in anaerobic environments. Annu Rev Microbiol 49:399–426.
5. Hungate RE. 1975. The rumen microbial ecosystem. Annual Review of Ecology and Systematics 6:39–66.
6. Weimer PJ. 1992. Cellulose degradation by ruminal microorganisms. Critical Reviews in Biotechnology 12:189–223.
7. Hungate RE. 1947. Studies on cellulose fermentation: III. The culture and isolation for cellulose-decomposing bacteria from the rumen of cattle. J Bacteriol 53:631–645.
8. Hungate RE. 1950. The anaerobic mesophilic cellulolytic bacteria. Bacteriol Rev 14:1–49.
9. Bryant MP, Doetsch RN. 1954. A study of actively cellulolytic rod-shaped bacteria of the bovine rumen. J Dairy Sci 37:1176–1183.
10. Montgomery L, Flesher B, Stahl D. 1988. Transfer of *Bacteroides succinogenes* (Hungate) to *Fibrobacter* gen. nov. as *Fibrobacter succinogenes* comb. nov. and description of *Fibrobacter intestinalis* sp. nov. International Journal of Systematic and Evolutionary Microbiology 38:430–435.
11. Lin C, Flesher B, Capman WC, Amann RI, Stahl DA. 1994. Taxon specific hybridization probes for fiber-digesting bacteria suggest novel gut-associated *Fibrobacter*. Systematic and Applied Microbiology 17:418–424.
12. Ilmberger N, Güllert S, Dannenberg J, Rabausch U, Torres J, Wemheuer B, Alawi M, Poehlein A, Chow J, Turaev D, Rattei T, Schmeisser C, Salomon J, Olsen PB, Daniel R, Grundhoff A, Borchert MS, Streit WR. 2014. A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. PLoS ONE 9:e106707.
13. Hongoh Y, Deevong P, Hattori S, Inoue T, Noda S, Noparatnaraporn N, Kudo T, Ohkuma M. 2006. Phylogenetic diversity, localization, and cell morphologies of members of the candidate phylum TG3 and a subphylum in the phylum Fibrobacteres, recently discovered bacterial groups dominant in termite guts. Appl Environ Microbiol 72:6780–6788.
14. McDonald JE, de Menezes AB, Allison HE, McCarthy AJ. 2009. Molecular biological detection and quantification of novel *Fibrobacter* populations in freshwater lakes. Appl Environ Microbiol 75:5148–5152.
15. McDonald JE, Lockhart RJ, Cox MJ, Allison HE, McCarthy AJ. 2008. Detection of novel *Fibrobacter* populations in landfill sites and determination of their relative abundance via quantitative PCR. Environ Microbiol 10:1310–1319.
16. Amann RI, Krumholz L, Stahl DA. 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. J Bacteriol 172:762–770.

17. Paster BJ, Ludwig W, Weisburg WG, Stackebrandt E, Hespell RB, Hahn CM, Reichenbach H, Stetter KO, Woese CR. 1985. A phylogenetic grouping of the *Bacteroides*, Cytophagas, and certain Flavobacteria. Systematic and Applied Microbiology 6:34–42.
18. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D, Brumm PJ. 2011. The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. PLoS ONE 6:e18814.
19. Abdul Rahman N, Parks DH, Vanwonterghem I, Morrison M, Tyson GW, Hugenholtz P. 2015. A phylogenomic analysis of the bacterial phylum Fibrobacteres. Front Microbiol 6:1469.
20. Montgomery L, Macy JM. 1982. Characterization of rat cecum cellulolytic bacteria. Appl Environ Microbiol 44:1435–1443.
21. Amann RI, Lin C, Key R, Montgomery L, Stahl DA. 1992. Diversity among *Fibrobacter* isolates: towards a phylogenetic classification. Systematic and Applied Microbiology 15:23–31.
22. Jewell KA, Scott JJ, Adams SM, Suen G. 2013. A phylogenetic analysis of the phylum Fibrobacteres. Syst Appl Microbiol 36:376–382.
23. Ransom-Jones E, Jones DL, Edwards A, McDonald JE. 2014. Distribution and diversity of members of the bacterial phylum Fibrobacteres in environments where cellulose degradation occurs. Syst Appl Microbiol 37:502–509.
24. Lin C, Stahl DA. 1995. Taxon-specific probes for the cellulolytic genus *Fibrobacter* reveal abundant and novel equine-associated populations. Appl Environ Microbiol 61:1348–1351.
25. Matsui H, Kato Y, Chikaraishi T, Moritani M, Ban-Tokuda T, Wakita M. 2010. Microbial diversity in ostrich ceca as revealed by 16S ribosomal RNA gene clone library and detection of novel *Fibrobacter* species. Anaerobe 16:83–93.
26. Sorokin DY, Gumerov VM, Rakitin AL, Beletsky AV, Damsté JSS, Muyzer G, Mardanov AV, Ravin NV. 2014. Genome analysis of *Chitinivibrio alkaliphilus* gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum Termite Group 3. Environ Microbiol 16:1549–1565.
27. Dehority BA. 1963. Isolation and characterization of several cellulolytic bacteria from in vitro rumen fermentations. Journal of Dairy Science 46:217–222.
28. Varel VH, Fryda SJ, Robinson IM. 1984. Cellulolytic bacteria from pig large intestine. Appl Environ Microbiol 47:219–221.
29. Stewart CS, Paniagua C, Dinsdale D, Cheng KJ, Garrow SH. 1981. Selective isolation and characteristics of *Bacteriodes succinogenes* from the rumen of a cow. Appl Environ Microbiol 41:504–510.
30. Dehority BA. 1965. Degradation and utilization of isolated hemicellulose by pure cultures of cellulolytic rumen bacteria. J Bacteriol 89:1515–1520.
31. Coen JA, Dehority BA. 1970. Degradation and utilization of hemicellulose from intact forages by pure cultures of rumen bacteria. Appl Microbiol 20:362–368.
32. Matte A, Forsberg CW, Verrinder Gibbins AM. 1992. Enzymes associated with metabolism of xylose and other pentoses by *Prevotella* (*Bacteroides*) *ruminicola* strains, *Selenomonas ruminantium* D, and *Fibrobacter succinogenes* S85. Can J Microbiol 38:370–376.

33. Dehority BA. 1971. Carbon dioxide requirement of various species of rumen bacteria. J Bacteriol 105:70–76.
34. Bryant MP, Robinson IM, Chu H. 1959. Observations on the nutrition of *Bacteroides succinogenes*—A ruminal cellulolytic bacterium. J Dairy Sci 42:1831–1847.
35. Bryant MP, Doetsch RN. 1954. Factors necessary for the growth of *Bacteroides succinogenes* in the volatile acid fraction of rumen fluid. Science 120:944–945.
36. Wegner GH, Foster EM. 1963. Incorporation of isobutyrate and valerate into cellular plasmalogen by *Bacteroides succinogenes*. J Bacteriol 85:53–61.
37. Dehority BA, Scott HW, Kowaluk P. 1967. Volatile fatty acid requirements of cellulolytic rumen bacteria. J Bacteriol 94:537–543.
38. Matheron C, Delort AM, Gaudet G, Forano E. 1998. In vivo 13C NMR study of glucose and cellobiose metabolism by four cellulolytic strains of the genus *Fibrobacter*. Biodegradation 9:451–461.
39. Franklund CV, Glass TL. 1987. Glucose uptake by the cellulolytic ruminal anaerobe *Bacteroides succinogenes*. J Bacteriol 169:500–506.
40. Maas LK, Glass TL. 1991. Cellobiose uptake by the cellulolytic ruminal anaerobe *Fibrobacter* (*Bacteroides*) *succinogenes*. Can J Microbiol 37:141–147.
41. Martin SA, Russell JB. 1986. Phosphoenolpyruvate-dependent phosphorylation of hexoses by ruminal bacteria: evidence for the phosphotransferase transport system. Appl Environ Microbiol 52:1348–1352.
42. Glass TL, Sherwood JS. 1994. Phosphorylation of glucose by a guanosine-5'-triphosphate (GTP)-dependent glucokinase in *Fibrobacter succinogenes* subsp. *succinogenes* S85. Arch Microbiol 162:180–186.
43. Wells JE, Russell JB, Shi Y, Weimer PJ. 1995. Cellodextrin efflux by the cellulolytic ruminal bacterium *Fibrobacter succinogenes* and its potential role in the growth of nonadherent bacteria. Appl Environ Microbiol 61:1757–1762.
44. Miller TL. 1978. The pathway of formation of acetate and succinate from pyruvate by *Bacteroides succinogenes*. Arch Microbiol 117:145–152.
45. Matheron C, Delort AM, Gaudet G, Forano E. 1997. Re-investigation of glucose metabolism in *Fibrobacter succinogenes*, using NMR spectroscopy and enzymatic assays. Evidence for pentose phosphates phosphoketolase and pyruvate formate lyase activities. Biochim Biophys Acta 1355:50–60.
46. Matheron C, Delort AM, Gaudet G, Liptaj T, Forano E. 1999. Interactions between carbon and nitrogen metabolism in *Fibrobacter succinogenes* S85: a 1H and 13C nuclear magnetic resonance and enzymatic study. Appl Environ Microbiol 65:1941–1948.
47. Meinhardt SW, Glass TL. 1994. NADH-linked fumarate reductase and NADH dehydrogenase activities in *Fibrobacter succinogenes*. Curr Microbiol 28:247–251.
48. Joyner AE, Baldwin RL. 1966. Enzymatic studies of pure cultures of rumen microorganisms. J Bacteriol 92:1321–1330.
49. Atasoglu C, Newbold CJ, Wallace RJ. 2001. Incorporation of [(15)N] ammonia by the cellulolytic ruminal bacteria *Fibrobacter succinogenes* BL2, *Ruminococcus albus* SY3, and *Ruminococcus flavefaciens* 17. Appl Environ Microbiol 67:2819–2822.
50. Bryant MP, Robinson IM. 1961. Studies on the nitrogen requirements of some ruminal cellulolytic bacteria. Appl Microbiol 9:96–103.
51. Dawson KA, Preziosi MC, Caldwell DR. 1979. Some effects of uncouplers and inhibitors on growth and electron transport in rumen bacteria. J Bacteriol 139:384–392.

52. Reddy CA, Bryant MP. 1977. Deoxyribonucleic acid base composition of certain species of the genus *Bacteroides*. Can J Microbiol 23:1252–1256.
53. Russell JB. 1987. Effect of extracellular pH on growth and proton motive force of *Bacteroides succinogenes*, a cellulolytic ruminal bacterium. Appl Environ Microbiol 53:2379–2383.
54. Russell JB, Dombrowski DB. 1980. Effect of pH on the efficiency of growth by pure cultures of rumen bacteria in continuous culture. Appl Environ Microbiol 39:604–610.
55. Wells JE, Russell JB. 1996. The effect of growth and starvation on the lysis of the ruminal cellulolytic bacterium *Fibrobacter succinogenes*. Appl Environ Microbiol 62:1342–1346.
56. Wells JE, Russell JB. 1994. The endogenous metabolism of *Fibrobacter succinogenes* and its relationship to cellobiose transport, viability and cellulose digestion. Appl Microbiol Biotechnol 41:471–476.
57. Gaudet G, Forano E, Dauphin G, Delort AM. 1992. Futile cycling of glycogen in *Fibrobacter succinogenes* as shown by in situ 1H-NMR and 13C-NMR investigation. Eur J Biochem 207:155–162.
58. Matheron C, Delort AM, Gaudet G, Forano E. 1996. Simultaneous but differential metabolism of glucose and cellobiose in *Fibrobacter succinogenes* cells, studied by in vivo 13C-NMR. Can J Microbiol 42:1091–1099.
59. Maglione G, Russell JB. 1997. The adverse effect of nitrogen limitation and excess-cellobiose on *Fibrobacter succinogenes* S85. Appl Microbiol Biotechnol 48:720–725.
60. Thomas S, Russell JB. 2004. The effect of cellobiose, glucose, and cellulose on the survival of *Fibrobacter succinogenes* A3C cultures grown under ammonia limitation. Curr Microbiol 48:219–223.
61. Matulova M, Delort AM, Nouaille R, Gaudet G, Forano E. 2001. Concurrent maltodextrin and cellodextrin synthesis by *Fibrobacter succinogenes* S85 as identified by 2D NMR spectroscopy. Eur J Biochem 268:3907–3915.
62. Nouaille R, Matulova M, Delort A-M, Forano E. 2004. Production of maltodextrin 1-phosphate by *Fibrobacter succinogenes* S85. FEBS Letters 576:226–230.
63. Nouaille R, Matulova M, Delort A-M, Forano E. 2005. Oligosaccharide synthesis in *Fibrobacter succinogenes* S85 and its modulation by the substrate. FEBS J 272:2416–2427.
64. Nouaille R, Matulova M, Pätoprstý V, Delort A-M, Forano E. 2009. Production of oligosaccharides and cellobionic acid by *Fibrobacter succinogenes* S85 growing on sugars, cellulose and wheat straw. Appl Microbiol Biotechnol 83:425–433.
65. Henderson G, Cox F, Ganesh S, Jonker A, Young W, Collaborators GRC, Abecia L, Angarita E, Aravena P, Arenas GN, Ariza C, Attwood GT, Avila JM, Avila-Stagno J, Bannink A, Barahona R, Batistotti M, Bertelsen MF, Brown-Kav A, Carvajal AM, Cersosimo L, Chaves AV, Church J, Clipson N, Cobos-Peralta MA, Cookson AL, Cravero S, Carballo OC, Crosley K, Cruz G, Cucchi MC, Barra R de la, Menezes ABD, Detmann E, Dieho K, Dijkstra J, Reis WLS dos, Dugan MER, Ebrahimi SH, Eythórsdóttir E, Fon FN, Fraga M, Franco F, Friedeman C, Fukuma N, Gagić D, Gangnat I, Grilli DJ, Guan LL, Miri VH, Hernandez-Sanabria E, Gomez AXI, Isah OA, Ishaq S, Jami E, Jelincic J, Kantanen J, Kelly WJ, Kim S-H, Klieve A, Kobayashi Y, Koike S, Kopecny J, Kristensen TN, Krizsan SJ, LaChance H, Lachman M, Lamberson WR, Lambie S, Lassen J, Leahy SC, Lee S-S, Leiber F, Lewis E, Lin B, Lira R, Lund P,

Macipe E, Mamuad LL, Mantovani HC, Marcoppido GA, Márquez C, Martin C, Martinez G, Martinez ME, Mayorga OL, McAllister TA, McSweeney C, Mestre L, Minnee E, Mitsumori M, Mizrahi I, Molina I, Muenger A, Muñoz C, Murovec B, Newbold J, Nsereko V, O'Donovan M, Okunade S, O'Neill B, Ospina S, Ouwerkerk D, Parra D, Pereira LGR, Pinares-Patiño C, Pope PB, Poulsen M, Rodehutscord M, Rodriguez T, Saito K, Sales F, Sauer C, Shingfield K, Shoji N, Simunek J, Stojanović-Radić Z, Stres B, Sun X, Swartz J, Tan ZL, Tapio I, Taxis TM, Tomkins N, Ungerfeld E, Valizadeh R, Adrichem P van, Hamme JV, Hoven WV, Waghorn G, Wallace RJ, Wang M, Waters SM, Keogh K, Witzig M, Wright A-DG, Yamano H, Yan T, Yáñez-Ruiz DR, Yeoman CJ, Zambrano R, Zeitz J, Zhou M, Zhou HW, Zou CX, Zunino P, 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.

66. Cheng K-J, Stewart CS, Dinsdale D, Costerton JW. 1984. Electron microscopy of bacteria involved in the digestion of plant cell walls. Animal Feed Science and Technology 10:93–120.
67. McDonald JE, Houghton JNI, Rooks DJ, Allison HE, McCarthy AJ. 2012. The microbial ecology of anaerobic cellulose degradation in municipal waste landfill sites: evidence of a role for fibrobacters. Environ Microbiol 14:1077–1087.
68. van Gylswyk NO, vander Toorn JJTK. 1986. Enumeration of *Bacteroides succinogenes* in the rumen of sheep fed maize-straw diets. FEMS Microbiol Ecol 2:205–209.
69. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. 2008. Evolution of mammals and their gut microbes. Science 320:1647–1651.
70. Macy JM, Farrand JR, Montgomery L. 1982. Cellulolytic and non-cellulolytic bacteria in rat gastrointestinal tracts. Appl Environ Microbiol 44:1428–1434.
71. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR. 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450:560–565.
72. Koike S, Pan J, Suzuki T, Takano T, Oshima C, Kobayashi Y, Tanaka K. 2004. Ruminal distribution of the cellulolytic bacterium *Fibrobacter succinogenes* in relation to its phylogenetic grouping. Animal Science Journal 75:417–422.
73. Fondevila M, Dehority BA. 1994. Degradation and utilization of forage hemicellulose by rumen bacteria, singly in coculture or added sequentially. J Appl Bacteriol 77:541–548.
74. Fondevila M, Dehority BA. 1996. Interactions between *Fibrobacter succinogenes*, *Prevotella ruminicola*, and *Ruminococcus flavefaciens* in the digestion of cellulose from forages. J Anim Sci 74:678–684.
75. 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*. Can J Microbiol 39:780–786.
76. Miron J, Ben-Ghedalia D. 1993. Digestion of structural polysaccharides of *Panicum* and vetch hays by the rumen bacterial strains *Fibrobacter succinogenes* BL2 and *Butyrivibrio fibrisolvens* D1. Appl Microbiol Biotechnol 39:756–759.

77. 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. J Appl Bacteriol 76:282–287.
78. Chen J, Weimer P. 2001. Competition among three predominant ruminal cellulolytic bacteria in the absence or presence of non-cellulolytic bacteria. Microbiology (Reading, Engl) 147:21–30.
79. Sawanon S, Koike S, Kobayashi Y. 2011. Evidence for the possible involvement of *Selenomonas ruminantium* in rumen fiber digestion. FEMS Microbiol Lett 325:170–179.
80. Scheifinger CC, Wolin MJ. 1973. Propionate formation from cellulose and soluble sugars by combined cultures of *Bacteroides succinogenes* and *Selenomonas ruminantium*. Appl Microbiol 26:789–795.
81. Stanton TB, Canale-Parola E. 1980. *Treponema bryantii* sp. nov., a rumen spirochete that interacts with cellulolytic bacteria. Arch Microbiol 127:145–156.
82. Weimer PJ. 1993. Effects of dilution rate and pH on the ruminal cellulolytic bacterium *Fibrobacter succinogenes* S85 in cellulose-fed continuous culture. Arch Microbiol 160:288–294.
83. Kudo H, Cheng KJ, Costerton JW. 1987. Interactions between *Treponema bryantii* and cellulolytic bacteria in the in vitro degradation of straw cellulose. Can J Microbiol 33:244–248.
84. Maglione G, Russell JB, Wilson DB. 1997. Kinetics of cellulose digestion by *Fibrobacter succinogenes* S85. Appl Environ Microbiol 63:665–669.
85. Kobayashi Y, Shinkai T, Koike S. 2008. Ecological and physiological characterization shows that *Fibrobacter succinogenes* is important in rumen fiber digestion - review. Folia Microbiol (Praha) 53:195–200.
86. Akin DE. 1980. Evaluation by electron microscopy and anaerobic culture of types of rumen bacteria associated with digestion of forage cell walls. Appl Environ Microbiol 39:242–252.
87. Shi Y, Odt CL, Weimer PJ. 1997. Competition for cellulose among three predominant ruminal cellulolytic bacteria under substrate-excess and substrate-limited conditions. Appl Environ Microbiol 63:734–742.
88. Mosoni P, Fonty G, Gouet P. 1997. Competition between ruminal cellulolytic bacteria for adhesion to cellulose. Curr Microbiol 35:44–47.
89. Shi Y, Weimer PJ. 1996. Utilization of individual cellodextrins by three predominant ruminal cellulolytic bacteria. Appl Environ Microbiol 62:1084–1088.
90. Latham MJ, Brooker BE, Pettipher GL, Harris PJ. 1978. Adhesion of *Bacteroides succinogenes* in pure culture and in the presence of *Ruminococcus flavefaciens* to cell walls in leaves of perennial ryegrass (*Lolium perenne*). Appl Environ Microbiol 35:1166–1173.
91. Bhat S, Wallace RJ, Orskov ER. 1990. Adhesion of cellulolytic ruminal bacteria to barley straw. Appl Environ Microbiol 56:2698–2703.
92. Odenyo AA, Mackie RI, Stahl DA, White BA. 1994. The use of 16S rRNA-targeted oligonucleotide probes to study competition between ruminal fibrolytic bacteria: pure-culture studies with cellulose and alkaline peroxide-treated wheat straw. Appl Environ Microbiol 60:3697–3703.

93. 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 Microbiol Ecol 92.
94. Koike S, Pan J, Kobayashi Y, Tanaka K. 2003. Kinetics of in sacco fiber-attachment of representative ruminal cellulolytic bacteria monitored by competitive PCR. J Dairy Sci 86:1429–1435.
95. Halliwell G, Bryant MP. 1963. The cellulolytic activity of pure strains of bacteria from the rumen of cattle. J Gen Microbiol 32:441–448.
96. Dehority BA, Scott HW. 1967. Extent of cellulose and hemicellulose digestion in various forages by pure cultures of rumen bacteria. J Dairy Sci 50:1136–1141.
97. Shinkai T, Ohji R, Matsumoto N, Kobayashi Y. 2009. Fibrolytic capabilities of ruminal bacterium *Fibrobacter succinogenes* in relation to its phylogenetic grouping. FEMS Microbiol Lett 294:183–190.
98. Joblin KN, Matsui H, Naylor GE, Ushida K. 2002. Degradation of fresh ryegrass by methanogenic co-cultures of ruminal fungi grown in the presence or absence of *Fibrobacter succinogenes*. Curr Microbiol 45:46–53.
99. Irvine HL, Stewart C s. 1991. Interactions between anaerobic cellulolytic bacteria and fungi in the presence of *Methanobrevibacter smithii*. Letters in Applied Microbiology 12:62–64.
100. Roger V, Grenet E, Jamot J, Bernalier A, Fonty G, Gouet P. 1992. Degradation of maize stem by two rumen fungal species, *Piromyces communis* and *Caecomyces communis*, in pure cultures or in association with cellulolytic bacteria. Reprod Nutr Dev 32:321–329.
101. Bernalier A, Fonty G, Bonnemoy F, Gouet P. 1992. Degradation and fermentation of cellulose by the rumen anaerobic fungi in axenic cultures or in association with cellulolytic bacteria. Curr Microbiol 25:143–148.
102. Weimer PJ. 1996. Why don't ruminal bacteria digest cellulose faster? J Dairy Sci 79:1496–1502.
103. Brumm P, Mead D, Boyum J, Drinkwater C, Deneke J, Gowda K, Stevenson D, Weimer P. 2011. Functional annotation of *Fibrobacter succinogenes* S85 carbohydrate active enzymes. Appl Biochem Biotechnol 163:649–657.
104. Béra-Maillet C, Ribot Y, Forano E. 2004. Fiber-degrading systems of different strains of the genus *Fibrobacter*. Appl Environ Microbiol 70:2172–2179.
105. Groleau D, Forsberg CW. 1981. Cellulolytic activity of the rumen bacterium *Bacteroides succinogenes*. Can J Microbiol 27:517–530.
106. Huang L, Forsberg CW. 1990. Cellulose digestion and cellulase regulation and distribution in *Fibrobacter succinogenes* subsp. *succinogenes* S85. Appl Environ Microbiol 56:1221–1228.
107. Béra-Maillet C, Gaudet G, Forano E. 2000. Endoglucanase activity and relative expression of glycoside hydrolase genes of *Fibrobacter succinogenes* S85 grown on different substrates. Biochim Biophys Acta 1543:77–85.
108. Gong J, Forsberg CW. 1989. Factors affecting adhesion of *Fibrobacter succinogenes* subsp. *succinogenes* S85 and adherence-defective mutants to cellulose. Appl Environ Microbiol 55:3039–3044.
109. Fields MW, Mallik S, Russell JB. 2000. *Fibrobacter succinogenes* S85 ferments ball-milled cellulose as fast as cellobiose until cellulose surface area is limiting. Appl Microbiol Biotechnol 54:570–574.

110. Roger V, Fonty G, Komisarczuk-Bony S, Gouet P. 1990. Effects of physicochemical factors on the adhesion to cellulose Avicel of the ruminal bacteria *Ruminococcus flavefaciens* and *Fibrobacter succinogenes* subsp. *succinogenes*. Appl Environ Microbiol 56:3081–3087.
111. Minato H, Suto T. 1978. Technique for fractionation of bacteria in rumen microbial ecosystem. J Gen Appl Microbiol 24:1–16.
112. Kudo H, Cheng KJ, Costerton JW. 1987. Electron microscopic study of the methylcellulose-mediated detachment of cellulolytic rumen bacteria from cellulose fibers. Can J Microbiol 33:267–272.
113. Miron J, Ben-Ghedalia D, Morrison M. 2001. Invited review: adhesion mechanisms of rumen cellulolytic bacteria. J Dairy Sci 84:1294–1309.
114. Raut MP, Karunakaran E, Mukherjee J, Biggs CA, Wright PC. 2015. Influence of substrates on the surface characteristics and membrane proteome of *Fibrobacter succinogenes* S85. PLOS ONE 10:e0141197.
115. Jun H-S, Qi M, Gong J, Egbosimba EE, Forsberg CW. 2007. Outer membrane proteins of *Fibrobacter succinogenes* with potential roles in adhesion to cellulose and in cellulose digestion. J Bacteriol 189:6806–6815.
116. Gong J, Egbosimba EE, Forsberg CW. 1996. Cellulose-binding proteins of *Fibrobacter succinogenes* and the possible role of a 180-kDa cellulose-binding glycoprotein in adhesion to cellulose. Can J Microbiol 42:453–460.
117. Burnet MC, Dohnalkova AC, Neumann AP, Lipton MS, Smith RD, Suen G, Callister SJ. 2015. Evaluating Models of Cellulose Degradation by *Fibrobacter succinogenes* S85. PLOS ONE 10:e0143809.
118. McGavin M, Forsberg CW. 1988. Isolation and characterization of endoglucanases 1 and 2 from *Bacteroides succinogenes* S85. J Bacteriol 170:2914–2922.
119. Blatch GL, Lässle M. 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays 21:932–939.
120. Ransom-Jones E, McCarthy AJ, Haldenby S, Doonan J, McDonald JE. 2017. Lignocellulose-degrading microbial communities in landfill sites represent a repository of unexplored biomass-degrading diversity. mSphere 2.
121. McGavin M, Lam J, Forsberg CW. 1990. Regulation and distribution of *Fibrobacter succinogenes* subsp. *succinogenes* S85 endoglucanases. Appl Environ Microbiol 56:1235–1244.
122. Huang L, McGavin M, Forsberg CW, Lam JS, Cheng KJ. 1990. Antigenic nature of the chloride-stimulated cellobiosidase and other cellulases of *Fibrobacter succinogenes* subsp. *succinogenes* S85 and related fresh isolates. Appl Environ Microbiol 56:1229–1234.
123. Malburg Jr. LM, Forsberg CW. 1993. *Fibrobacter succinogenes* S85 possesses at least nine distinct glucanase genes. Can J Microbiol 39:882–891.
124. Huang L, Forsberg CW. 1987. Isolation of a cellodextrinase from *Bacteroides succinogenes*. Appl Environ Microbiol 53:1034–1041.
125. Huang L, Forsberg CW, Thomas DY. 1988. Purification and characterization of a chloride-stimulated cellobiosidase from *Bacteroides succinogenes* S85. J Bacteriol 170:2923–2932.

126. Huang L, Forsberg CW. 1988. Purification and comparison of the periplasmic and extracellular forms of the cellodextrinase from *Bacteroides succinogenes*. Appl Environ Microbiol 54:1488–1493.
127. Qi M, Jun H-S, Forsberg CW. 2008. Cel9D, an atypical 1,4-beta-D-glucan glucohydrolase from *Fibrobacter succinogenes*: characteristics, catalytic residues, and synergistic interactions with other cellulases. J Bacteriol 190:1976–1984.
128. Buchanan CJ, Mitchell WJ. 1992. Two beta-glucosidase activities in *Fibrobacter succinogenes* S85. J Appl Bacteriol 73:243–250.
129. Miron J, Yokoyama MT, Lamed R. 1989. Bacterial cell surface structures involved in lucerne cell wall degradation by pure cultures of cellulolytic rumen bacteria. Appl Microbiol Biotechnol 32:218–222.
130. Qi M, Jun H-S, Forsberg CW. 2007. Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. Appl Environ Microbiol 73:6098–6105.
131. Arntzen MØ, Várnai A, Mackie RI, Eijsink VGH, Pope PB. 2017. Outer membrane vesicles from *Fibrobacter succinogenes* S85 contain an array of carbohydrate-active enzymes with versatile polysaccharide-degrading capacity. Environ Microbiol 19:2701–2714.
132. Forsberg CW, Beveridge TJ, Hellstrom A. 1981. Cellulase and xylanase release from *Bacteroides succinogenes* and its importance in the rumen environment. Appl Environ Microbiol 42:886–896.
133. Gaudet G, Gaillard B. 1987. Vesicle formation and cellulose degradation in *Bacteroides succinogenes* cultures: ultrastructural aspects. Arch Microbiol 148:150–154.
134. Matulova M, Nouaille R, Capek P, Péan M, Forano E, Delort A-M. 2005. Degradation of wheat straw by *Fibrobacter succinogenes* S85: a liquid- and solid-state nuclear magnetic resonance study. Appl Environ Microbiol 71:1247–1253.
135. McDermid KP, Mackenzie CR, Forsberg CW. 1990. Esterase activities of *Fibrobacter succinogenes* subsp. *succinogenes* S85. Appl Environ Microbiol 56:127–132.
136. McDermid KP, Forsberg CW, MacKenzie CR. 1990. Purification and properties of an acetylxylan esterase from *Fibrobacter succinogenes* S85. Appl Environ Microbiol 56:3805–3810.
137. Kam DK, Jun H-S, Ha JK, Inglis GD, Forsberg CW. 2005. Characteristics of adjacent family 6 acetylxylan esterases from *Fibrobacter succinogenes* and the interaction with the Xyn10E xylanase in hydrolysis of acetylated xylan. Can J Microbiol 51:821–832.
138. Yoshida S, Mackie RI, Cann IKO. 2010. Biochemical and domain analyses of FSUAxe6B, a modular acetyl xylan esterase, identify a unique carbohydrate binding module in *Fibrobacter succinogenes* S85. J Bacteriol 192:483–493.
139. Jun HS, Ha JK, Malburg LM, Verrinder GAM, Forsberg CW. 2003. Characteristics of a cluster of xylanase genes in *Fibrobacter succinogenes* S85. Can J Microbiol 49:171–180.
140. Malburg LM, Smith DC, Schellhorn HE, Forsberg CW. 1993. *Fibrobacter succinogenes* S85 has multiple xylanase genes. Journal of Applied Bacteriology 75:564–573.

Chapter 2: *Fibrobacter* communities in the gastrointestinal tracts of diverse herbivores

Authors and contributors:

Anthony Neumann: Designed the experiments, performed the experiments, analyzed the data, and wrote the manuscript

Caroline McCormick: Performed the experiments, and edited the manuscript

Dr. Garret Suen: Designed the experiments, and edited the manuscript

Adapted from:

Neumann, A. P., McCormick, C. A., and Suen, G. *Fibrobacter* communities in the gastrointestinal tracts of diverse hindgut-fermenting herbivores are distinct from those of the rumen. Environ Microbiol. 2017 Sep;19(9):3768-3783.

Funding:

This work was supported by a traineeship from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number T32AI55397 to APN, a UW-Madison Hilldale Undergraduate Research fellowship to CAM, and a US Department of Energy Biological and Environmental Research Early Career Research Program Award DE-SC0008104 to GS. The funders had no input in the design, execution, or interpretation of this work.

2.1 Abstract

Herbivores engage in symbioses with gut microbiota enabling them to digest their plant-based diet. Efforts to understand these microbial communities has advanced significantly with the advent of culture-independent, sequence-based analyses. One challenge with this approach is the difficulty in testing hypotheses from these datasets due to a lack of cultured representatives for many of the taxa in these communities. This includes the Fibrobacteres, a bacterial phylum that contains only a handful of cultured isolates primarily in the genus *Fibrobacter*. Culture-independent investigations have identified *Fibrobacter* populations in the gastrointestinal tracts of diverse animals, particularly hindgut-fermenting herbivores, but their physiology remains unknown due to a paucity of representative axenic cultures. Here, we developed a novel isolation method and complemented it with 16S rRNA gene sequencing on samples collected from phylogenetically diverse herbivores. We isolated 45 new strains from 11 different hosts and identified 10 discrete phylotypes, including 4 previously unrepresented by cultures. Our 16S rRNA analysis indicated that most isolates were representative of the numerically dominant populations in their respective samples, confirming previous reports of different *Fibrobacter* populations in the rumen, compared to the hindgut of horses. Phenotypic characterization of these strains indicated that, while all could grow on crystalline cellulose, none exhibited growth on xylan. In addition to providing novel ecological and evolutionary insights into an important, but poorly understood phylum in the herbivore gut, this work significantly expands the diversity of *Fibrobacter* cultures available for study, and further describes a new generalizable approach for culturing isolates of anaerobic fiber degraders.

2.2 Introduction

Herbivores depend on microorganisms living in their gastrointestinal tract for the efficient digestion of their fiber-rich diet (1, 2). The plant cell wall polysaccharides (PCWP) that constitute the bulk of this fiber represent a potentially plentiful source of carbon that is highly resistant to enzymatic decomposition (3). Despite this recalcitrance, certain microbial taxa have evolved mechanisms to extract the sugars from these structural polysaccharides in order to ferment them for energy (4, 5). By maintaining a community of these microbes in their gut, herbivores gain access to the chemical energy in their diet through absorption of the byproducts of this microbial fermentation. Animals engaged in this symbiosis over extended periods of evolutionary time have optimized their digestive physiology through enlargements in either the foregut or hindgut, which increase the capacity and retention time of feed through the gastrointestinal tract (6). Although microbial consortia responsible for fiber digestion in foregut fermenters have been well characterized via studies of the rumen (7), less is known regarding the microbial taxa that perform this function in hindgut-fermenting herbivores.

Plant cell walls are primarily composed of cellulose, hemicellulose, pectin, and lignin (8). Cellulose is the most abundant of these PCWP, but its crystalline structure makes it among the most difficult to hydrolyze (9). Therefore, microorganisms that can efficiently deconstruct cellulose in the herbivore gut are critically important for optimal digestion. In the rumen, cellulose degradation is facilitated, in part, via the efforts of certain species of bacteria, including members of the genus *Fibrobacter* (7, 10–12). Representatives were first isolated and described by Hungate during studies of cellulose digestion in the bovine rumen (13). Although they were originally classified as *Bacteroides*, phylogenetic analyses have subsequently established their membership to a deeply rooted clade within the Bacteria designated as the phylum Fibrobacteres

(14, 15). The type strain, *Fibrobacter succinogenes* S85, has been extensively investigated and is among the most actively cellulolytic of all strictly anaerobic, mesophilic bacteria known (16–19). *F. succinogenes* S85 ferments the breakdown products of cellulose to primarily succinic acid and lesser amounts of acetic acid. Production of these fermentation products, along with an apparent inability to ferment pentoses, appears to be a conserved feature among *Fibrobacter* spp. (20, 21).

Since their original isolation from the rumen, culture-independent studies have suggested the presence of *Fibrobacter* populations in diverse herbivore gastrointestinal tracts (22–24). However, cultured representatives from most of these hosts are rare, and as a result, our understanding of their physiology is limited. Currently, there are only two formally described species for the genus *Fibrobacter*: *F. succinogenes* and *F. intestinalis* (15, 20). Phylogenetic analysis based on full-length 16S rRNA gene sequences of all of the available isolates from these two species groups indicates that there are at least four distinct phylogenetic lineages of *F. succinogenes* (20, 25), but culture-independent analyses suggest that this represents only a fraction of the true diversity in the *F. succinogenes* group (26, 27). Similarly, poor cultural representation almost certainly exists for *F. intestinalis*, since fewer representative isolates have been described. Moreover, while the phylogenetic distance between *F. succinogenes* and *F. intestinalis* is satisfactory for segregation into separate genera, sufficient phenotypic differentiation has not yet been established to justify reclassification (20).

The aim of this investigation was to gain insights into the ecology of *Fibrobacter* spp. in the gastrointestinal tracts of herbivores, particularly hindgut-fermenters. To achieve this goal, we complemented 16S rRNA gene amplicon sequencing with culturing and isolation, an approach that addresses some of the limitations associated with a solely culture-independent

analysis (28). We chose to focus on hindgut-fermenting herbivores because the bacterial communities involved in fiber-degradation in the gastrointestinal tracts of these hosts have generally been understudied, relative to analogous bacteria in ruminant livestock. We hypothesized that hindgut-fermenting herbivores would harbor phylogenetically distinct *Fibrobacter* populations in their gastrointestinal tracts, compared to those that have been observed in the rumen, but that their niche as cellulose-degrading specialists would be conserved. Here, we describe the isolation and ecological distribution of 45 novel *Fibrobacter* isolates, including four previously uncharacterized phylotypes, and further introduce a novel strategy for isolating these bacteria. Our results provide strong evidence that several relatively abundant *Fibrobacter* phylotypes found in hindgut fermenters are phylogenetically distinct from those typically observed in the rumen.

2.3 Results

A total of 95 samples, collected from animals housed at farms, university facilities, and two zoos in Wisconsin, USA, were examined for this study. The samples included 78 fresh fecal samples, 14 samples of rumen contents, and three samples of cecal contents. Samples were collected from 23 different animal hosts representing seven orders of Mammals: Artiodactyla (even-toed ungulates), Perissodactyla (odd-toed ungulates), Rodentia, Primates, Carnivora, Proboscidea, and Pilosa. Samples from one bird (ostrich) and one reptile (Aldabra tortoise) were also examined. These hosts exhibited a primarily herbivorous dietary lifestyle (n = 84), although a small number of samples from omnivores (n = 8) and carnivores (n = 3) were also investigated for comparison. Most of the samples represented hosts with a gastrointestinal anatomy specialized for either hindgut-fermentation (n = 61) or foregut-fermentation (n = 29) of PCWP.

Isolation of novel *Fibrobacter* phylotypes

A total of 45 axenic cultures of *Fibrobacter* from 41 of the 95 (43.2%) samples were recovered after all isolation attempts (Table 2.1). Isolates were recovered from 11 different host species including: Holstein cattle (n = 17), horse (n = 9), pig (n = 4), rhinoceros (n =4), tapir (n = 3), capybara (n = 2), sheep (n = 2), colobus monkey (n = 1), elephant (n = 1), ostrich (n = 1), and rhesus monkey (n = 1). Microscopic examination as well as the recovery of single 16S rRNA gene sequences, using the universal bacterial primers 27F and 1492R (29), from cultures grown in rich media supported the conclusion that these 45 cultures are axenic and are members of the phylum Fibrobacteres, with the closest cultured relatives in the genus *Fibrobacter*.

A maximum likelihood phylogeny showing the relationships among these 45 *Fibrobacter* strains was inferred from the near-full length 16S rRNA gene sequences (Figure 2.1). All but one of the 45 strains isolated in this study segregated into eight discrete and well resolved (≥ 98% bootstrap support) lineages. One strain isolated from capybara feces, UWP2, did not cluster with any of the others, suggesting a ninth distinct lineage. Additionally, none of the 45 strains clustered closely with *Fibrobacter succinogenes* strain HM2, the type strain for *F. succinogenes* subsp. *elongatus* (14), indicating at least ten discrete lineages represented by *Fibrobacter* isolates described to date. The phylogeny supported the placement of 39 of the 45 strains, representing seven lineages including strain HM2, in the *Fibrobacter succinogenes* species clade with high confidence (96% bootstrap support). The remaining six strains fell into three discrete lineages, two of which also contained previously described isolates of *F. intestinalis* (20). Based on these results, and following the nomenclature previously reported, these ten lineages are designated Fs I through Fs VII, for the seven lineages of the *F. succinogenes* clade, and Fi I through Fi III, for

Table 2.1: Sources of *Fibrobacter* isolates

Host Species	Sample Type	# of Isolates	# of Animals	# of sample sites	# of phylotypes
Cow (cattle)	R	17	14	3	2
Horse	F	9	8	4	2
Pig	F, C	4	4	2	3
Rhinoceros	F	4	2	2	2
Tapir	F	3	2	2	3
Capybara	F	2	unknown*	1	2
Sheep	R	2	2	1	1
Colobus monkey	F	1	unknown*	1	1
Elephant	F	1	1	1	1
Ostrich	F	1	1	1	1
Rhesus monkey	F	1	1	1	1

R = rumen contents, F = feces, C = cecal contents
*group sample

F. intestinalis and its closest relatives (Figure 2.1). Cultured representatives for three of the six *F. succinogenes* lineages represented by strains isolated in this study, Fs V, Fs VI, and Fs VII, have not been reported previously. Moreover, the strains UWRM and UWS4 are the first cultured representatives of Fi III, and although their closest previously cultured representative is *F. intestinalis* strain JG1, they share only 94% nucleotide identity with this strain across the corresponding 1,262 base pairs of 16S rRNA gene sequence examined. Clustering using mothur [(30), average neighbor, cutoff = 0.03] of the near-full length 16S rRNA gene sequences from the 45 *Fibrobacter* strains isolated in this study, along with the sequence from *F. succinogenes* strain HM2, identified ten distinct operational taxonomic units (OTUs) or phylotypes. The grouping of the strains into OTUs was in absolute agreement with the grouping observed in the maximum likelihood phylogeny, providing further support for the ten distinct *Fibrobacter* phylotypes and their member strains.

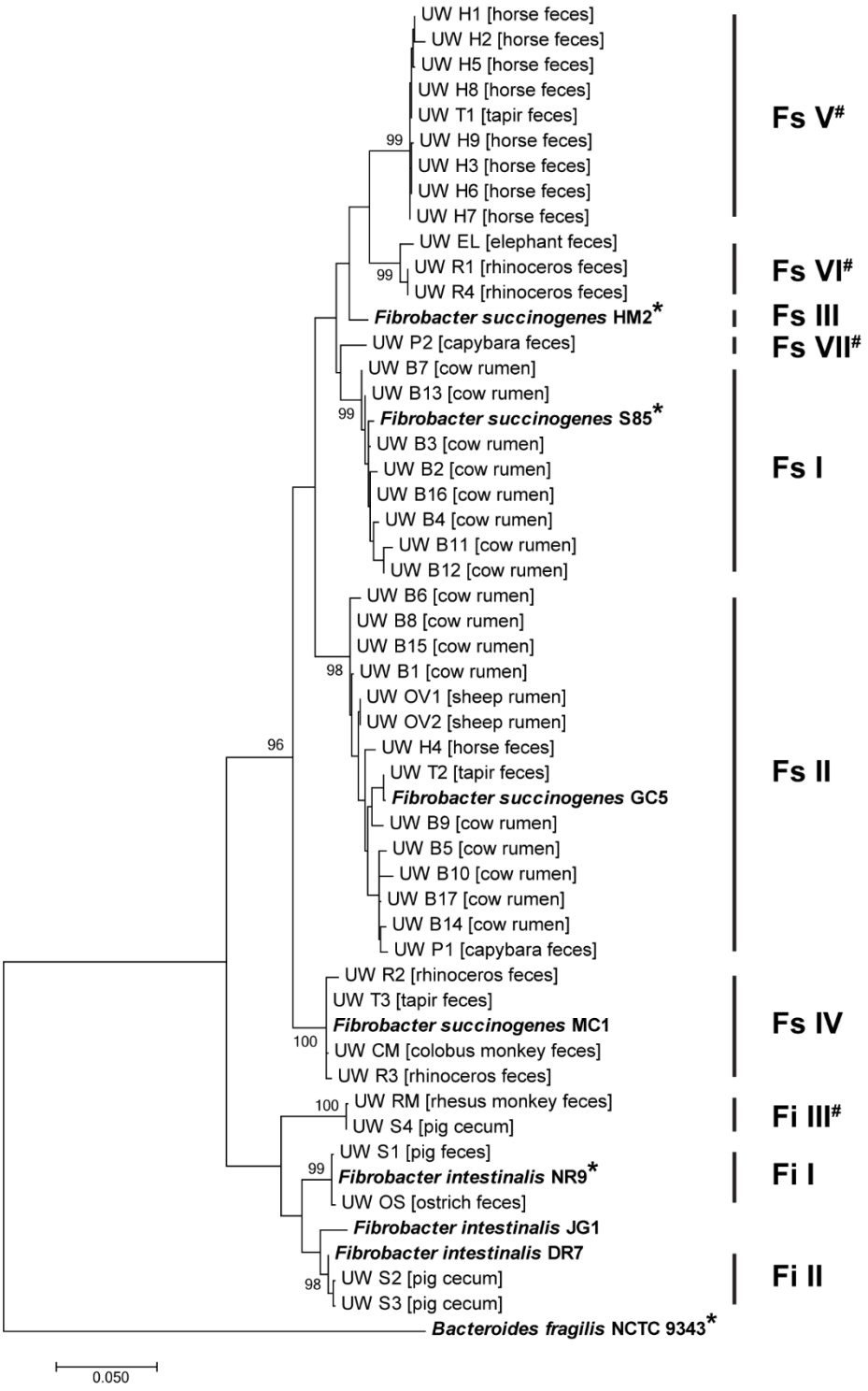


Figure 2.1: 16S rRNA gene maximum likelihood inferred phylogeny of *Fibrobacter* isolates. Reference sequences for previously described *Fibrobacter* strains are shown in bold with the type strains marked with an asterisk (*). *Fibrobacter* phylotypes are displayed on the far right, novel phylotypes are denoted with a hash (#). The *Bacteroides fragilis* type strain NCTC 9343 was included as an outgroup.

Culture-independent analysis of bacterial communities

Total bacterial community profiling by 16S rRNA gene amplicon sequencing was performed in order to determine whether the *Fibrobacter* isolates we obtained were representative of the abundant *Fibrobacter* populations in their respective samples, as well as to provide additional ecological context for the *Fibrobacter* phylotypes. Data for 83 of the samples used for culturing was obtained by Illumina sequencing of the V4 region (31). A scatterplot of the two-dimensional NMDS analysis based on Bray-Curtis dissimilarities among the total bacterial communities from these samples showed general clustering by host and, more broadly, by host phylogeny, with hosts from the same taxonomic order typically placed within similar coordinates of the Cartesian plane (Figure 2.2). Most samples were dominated by the phyla Firmicutes and Bacteroidetes. Other commonly observed phyla in gastrointestinal or fecal samples included: Spirochaetes, Verrucomicrobia, Proteobacteria, and Tenericutes. Classifiable genera that were commonly observed across samples include: *Prevotella*, *Oscillospira*, *Arcobacter*, *Ruminococcus*, and *Treponema*.

Samples were considered positive for the presence of Fibrobacteres if the relative abundance of sequences classified to this phylum was greater than 0.01%. Overall, 67 of the 83 samples analyzed by 16S rRNA gene amplicon sequencing were positive for Fibrobacteres sequences, with a median percent relative abundance among these samples of 0.38% (Figure 2.3). Most sequences (96.83%) that classified to the phylum Fibrobacteres were also classified to the genus *Fibrobacter*. All rumen samples (Holstein cows and sheep) and fecal samples from odd-toed ungulates (horses, tapirs, and rhinoceros) were positive for Fibrobacteres. Rumen samples had a median percent relative abundance of Fibrobacteres sequences of 1.03%, while

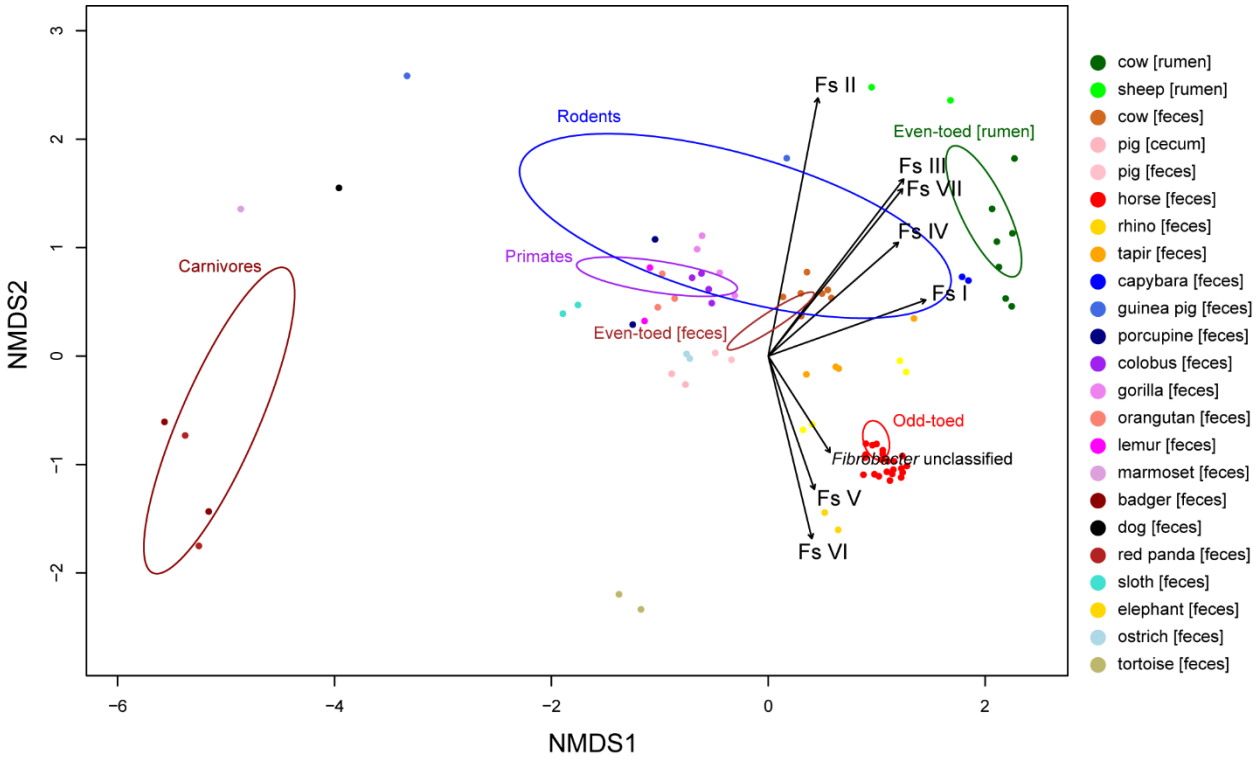


Figure 2.2: NMDS ordination plot of Bray-Curtis dissimilarity among the total bacterial communities. Total bacterial communities are represented by dots colored according to host of origin. Sets of samples corresponding to specific host taxonomic orders are plotted as open standard error ellipses (95% confidence interval), and colored/labeled accordingly. Vectors representing sequence counts for *Fibrobacter* phylotypes showing a possible association ($P < 0.2$) with a certain area of the ordination are shown and labeled accordingly.

fecal samples from odd-toed ungulates had an overall median of 0.41%. Less abundant Fibrobacteres populations were observed in fecal/cecal samples from even-toed ungulates (7 of 11 positive, median = 0.05%) and fecal samples from primates (11 of 15 positive, median = 0.05%). None of the fecal samples from carnivores (n = 5) met the criteria for positive detection of Fibrobacteres populations. The most extensively sampled hosts, rumen samples from Holstein cows (n = 7) and fecal samples from horses (n = 22), had similar median percent relative abundance of Fibrobacteres sequences of 0.65% and 0.54%, respectively (Figure 2.3). Relatively abundant Fibrobacteres populations were also observed for the limited number of

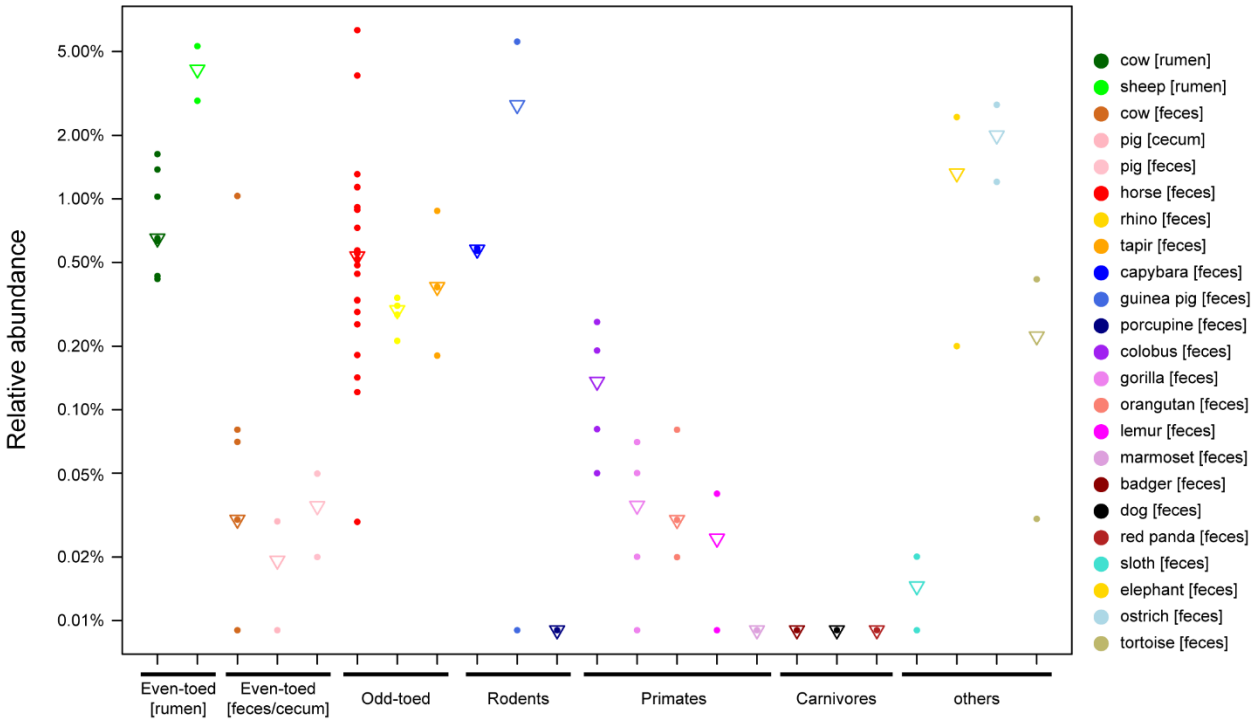


Figure 2.3: Percent relative abundance of Fibrobacteres sequences in samples by host. Each dot represents an individual sample, colored according to host of origin, and plotted on a logarithmic scale (y-axis). Median percent relative abundances for each host are plotted with open triangles. Samples are arranged by host taxonomic order, which are labeled along the x-axis.

fecal samples collected from the ostrich (2 of 2 positive, median = 2.00%), elephants (2 of 2 positive, median = 1.32%), and capybaras (2 of 2 positive, median = 0.57%).

Ecological differences among *Fibrobacter* phylotypes

We further analyzed those 16S rRNA amplicon sequences classified to the phylum Fibrobacteres by comparing them against the 16S rRNA gene sequences obtained from our 45 cultured strains and *F. succinogenes* subsp. *elongatus* strain HM2, to determine if they could be classified as belonging to any of the ten identified *Fibrobacter* phylotypes. The *Fibrobacter* phylotypes were then tested for associations with the dimensions of the NMDS ordination of Bray-Curtis dissimilarities among the total bacterial communities to identify possible

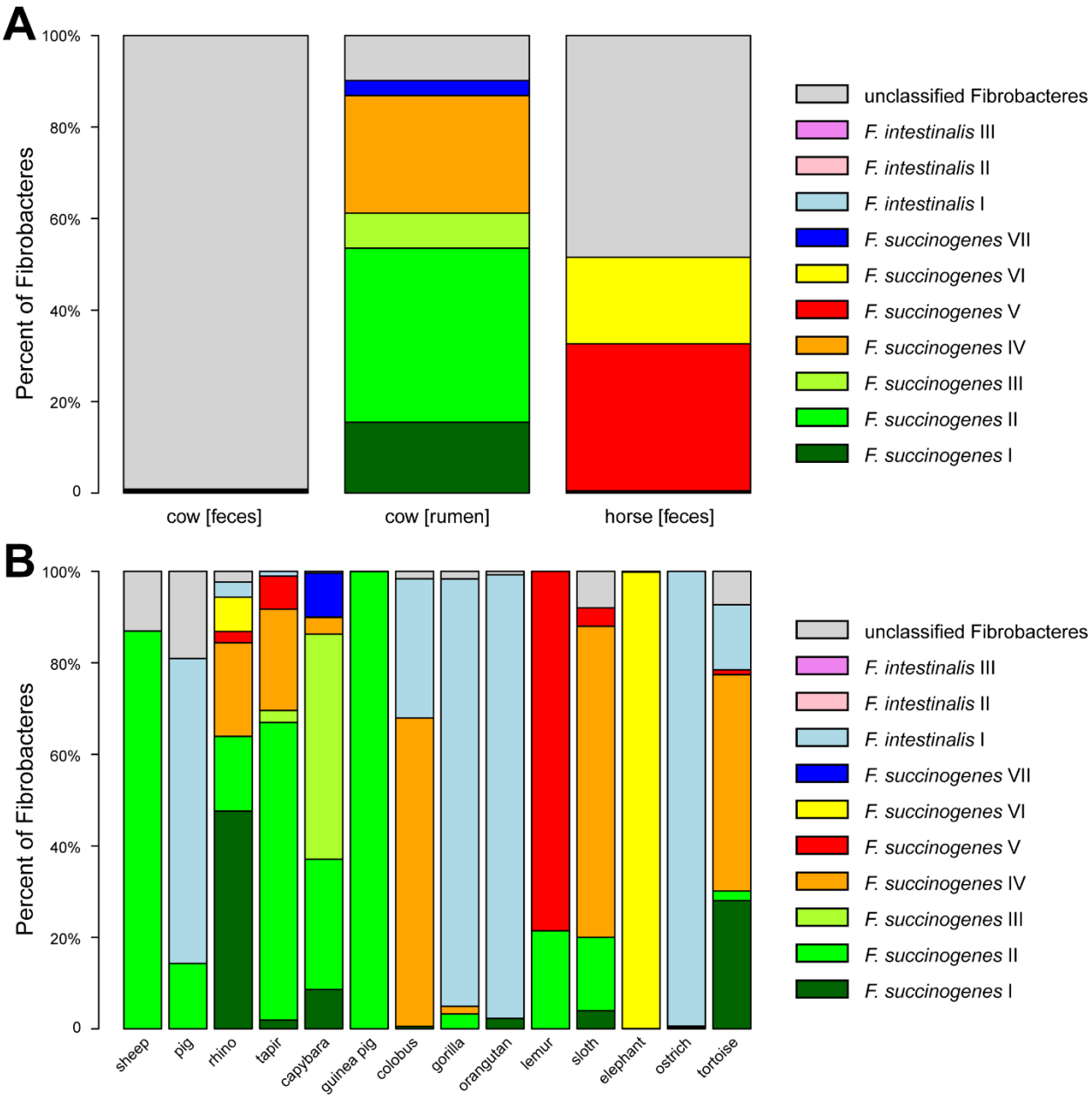


Figure 2.4: Fraction of Fibrobacteres sequences assigned to a specific *Fibrobacter* phylotype by host. (A) Distribution of *Fibrobacter* phylotypes in samples from Holstein cows and horses. (B) Distribution of *Fibrobacter* phylotypes in other hosts for which Fibrobacteres sequences were observed.

relationships with total bacterial communities from certain hosts. Vectors representing sequence counts for *Fibrobacter* phylotypes showing a possible association ($P < 0.2$) with a certain area of the ordination are shown in Figure 2.2. Vectors for the phylotypes Fs I, Fs II, Fs III, Fs IV, and

Fs VII were associated with an area of the plot that included total bacterial communities from rumen samples. In contrast, the vectors corresponding to sequence counts for *Fibrobacter* phylotypes Fs V and Fs VI, as well as a vector representing *Fibrobacter* sequences that did not classify to any of the ten phylotypes, were directed toward the region where total bacterial communities from the feces of odd-toed ungulates and elephants were plotted.

The fraction of 16S rRNA Fibrobacteres amplicon sequences belonging to the different *Fibrobacter* phylotypes for the most extensively sampled hosts in this study, Holstein cows (n = 7) and horses (n = 22), are shown in Figure 2.4A. Almost all (99.17%) of the Fibrobacteres sequences recovered from cow fecal samples could not be classified to any of the ten *Fibrobacter* phylotypes represented by the strains isolated in this study. In contrast, only 8.97% of the Fibrobacteres sequences recovered from cow rumen samples were not represented by one of the ten phylotypes. Five *Fibrobacter* phylotypes were consistently observed in cow rumen samples: Fs I, Fs II, Fs III, Fs IV, and Fs VII. None of the Fibrobacteres sequences recovered from rumen samples were classified as Fs V, Fs VI, or any of the three *F. intestinalis* phylotypes. The distribution of *Fibrobacter* phylotypes in horse fecal samples was distinct from what was observed for cow fecal samples. A total of 50.57% of all of the Fibrobacteres sequences from horse fecal samples could not be classified to any of the ten *Fibrobacter* phylotypes. However, sequences classified as either *Fibrobacter* phylotype Fs V or Fs VI were commonly observed in horse fecal samples, occurring in 72.73% and 81.82% of the samples examined, respectively. Sequences classified as Fs II, Fs III, and Fs IV were rarely observed in horse fecal samples, whereas phylotypes Fs I, Fs VII, Fi I, Fi II, and Fi III were not observed at all. The fraction of Fibrobacteres sequences classified to the *Fibrobacter* phylotypes in other hosts for which Fibrobacteres sequences were detected is shown in Figure 2.4B. Phylotype Fs II exhibited the

widest host range, with sequences detected in 11 different animal hosts. Sequences related to the *F. intestinalis* type strain, phylotype Fi I, dominated the Fibrobacteres community in pigs, gorillas, orangutans, and the ostrich. Sequences classified as Fi II or Fi III were rarely, if ever, observed.

Fibrobacter phylotypes in the Global Rumen Census

16S rRNA gene amplicon sequence data from the Global Rumen Census (GRC) was analyzed (32) in order to determine whether the absence of *Fibrobacter* phylotypes Fs V and Fs VI in rumen samples extended beyond the nine individual rumen samples investigated in our study. We classified 96,331 sequences out of 3,601,905 total bacterial sequences in the GRC to the phylum Fibrobacteres (2.67%). These Fibrobacteres sequences were then classified to our defined *Fibrobacter* phylotypes. Figure 2.5 shows the fraction of GRC Fibrobacteres sequences classified to our *Fibrobacter* phylotypes for each of eight different ruminant hosts. Overall, 89.02% of the GRC Fibrobacteres sequences could be classified to one of the ten *Fibrobacter* phylotypes. As was observed for rumen samples collected and analyzed for this study, sequences classified to phylotypes Fs I, Fs II, Fs III, and Fs IV were common in rumen samples analyzed for the GRC. Of these, Fs I and Fs II typically dominated the Fibrobacteres community, accounting for more than 60% of the Fibrobacteres sequences, on average, regardless of host (except in giraffes), as well as 72.79% of all Fibrobacteres sequences from the GRC. Lesser amounts of *Fibrobacter* phylotypes Fs VII, Fi II, and Fi III were also observed, including a moderate percentage of the Fibrobacteres sequences from giraffe rumen samples being classified as Fi III. Despite the substantial number of Fibrobacteres sequences recovered

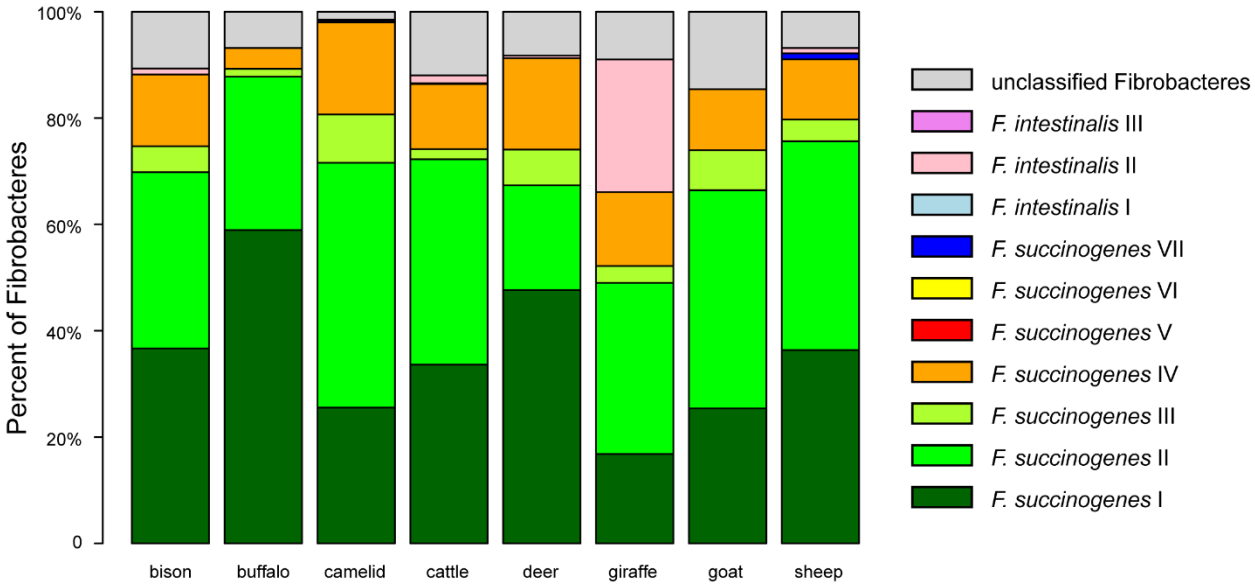


Figure 2.5: Fraction of Fibrobacteres sequences from the Global Rumen Census (32) assigned to a specific *Fibrobacter* phylotype by ruminant host.

and analyzed from the GRC dataset, none of the sequences were classified to the *Fibrobacter* phylotypes Fs V, Fs VI, or Fi I.

Ecology of unclassified Fibrobacteres phylotypes

An OTU-based analysis of the Fibrobacteres sequences identified among the 16S rRNA gene amplicon data generated in this study was performed in order to gain additional insights into the large proportions of unclassified Fibrobacteres sequences observed in fecal samples from Holstein cows and horses. After clustering Fibrobacteres sequences into OTUs (average neighbor, cutoff = 0.03), representative sequences were used to construct a maximum likelihood neighbor-joining tree of those OTUs observed in more than one sample as well as at a percent relative abundance of at least 0.1% in at least one sample (Figure 2.6). Of these 13 commonly observed Fibrobacteres OTUs, six represent phylogenetic lineages not represented by the ten phylotypes described in this study (Table 2.2). Two of the unrepresented Fibrobacteres OTUs,

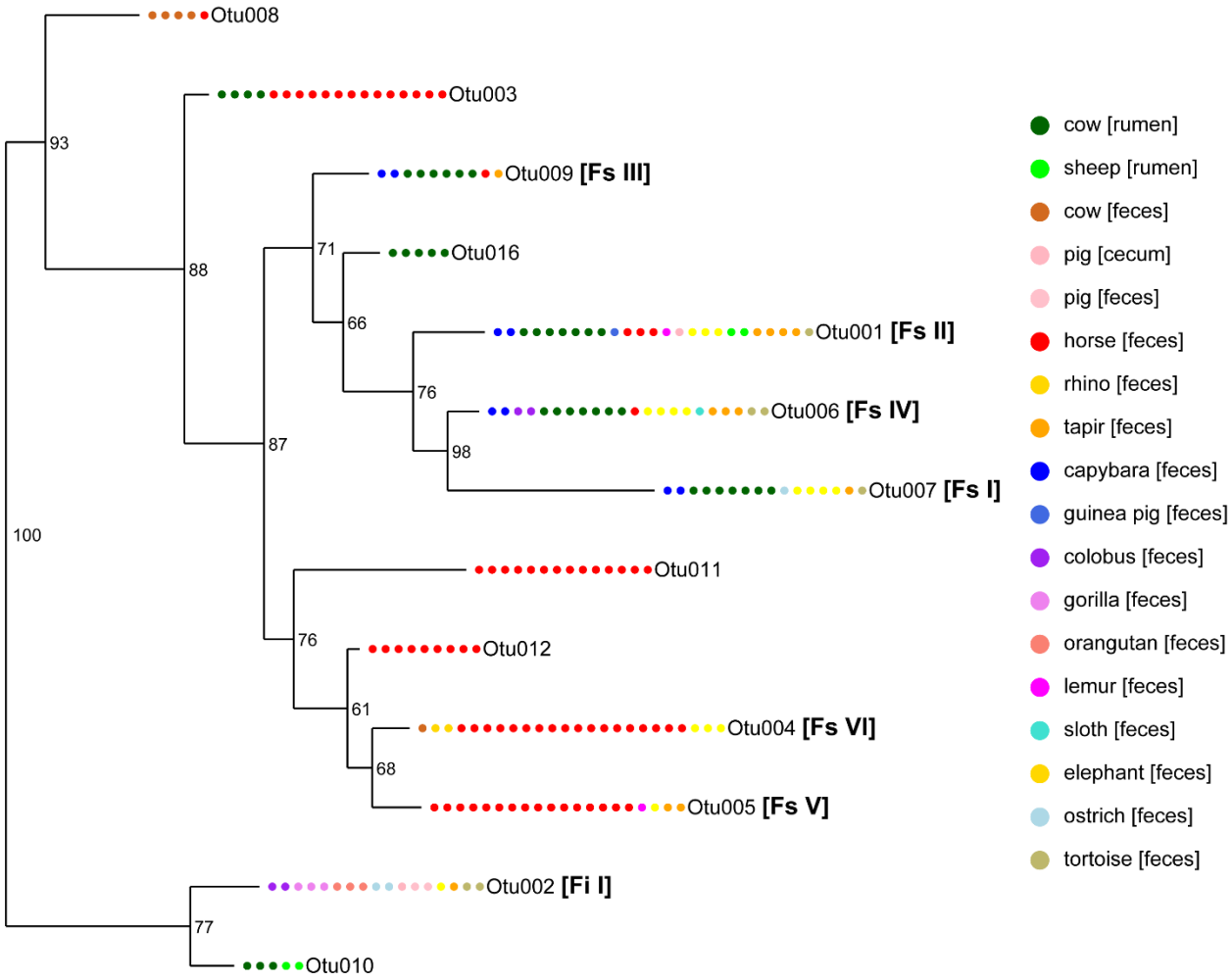


Figure 2.6: Phylogenetic relationships among common Fibrobacteres OTUs. A maximum likelihood neighbor-joining tree was inferred from the aligned nucleotide sequences representing commonly observed Fibrobacteres OTUs in the culture-independent analysis of total bacterial communities. Common OTUs were defined as those that were observed in more than one sample along with having a percent relative abundance of at least 0.1% in at least one sample. Nodes are labeled with the level of bootstrap support for their respective clades. Branch tips are labeled with dots, colored according to host of origin, indicating the occurrence of that particular OTU in a given microbiota sample, the OTU ID, and the corresponding *Fibrobacter* phylotype designation when appropriate (see Table 2.2).

Otu011 and Otu012, are close relatives of the cultured phylotypes Fs V and Fs VI and exhibit a similar pattern of ecological distribution being solely observed in horse fecal samples. Our OTU-based analysis also identified three distinct phylogenetic lineages located outside of the *F. succinogenes* group. Otu010 is likely representative of a separate phylotype within the *F.*

Table 2.2: 16S rRNA sequence similarity values for common Fibrobacteres OTUs

OTU	Closest reference	Identity	Length	Mismatch	Gap	e-value
Otu001	*F. succinogenes* GC5 [Fs II]	100%	253	0	0	1.64E-134
Otu002	*F. intestinalis* NR9* [Fi I]	100%	253	0	0	1.64E-134
Otu003	UW H9	96.84%	253	8	0	3.60E-121
Otu004	UW R1 [Fs VI]	99.61%	253	1	0	7.62E-133
Otu005	UW H9 [Fs V]	100%	253	0	0	1.64E-134
Otu006	UW T3 [Fs IV]	100%	253	0	0	1.64E-134
Otu007	*F. succinogenes* S85* [Fs I]	100%	253	0	0	1.64E-134
Otu008	UW P2	93.33%	255	14	3	3.67E-106
Otu009	*F. succinogenes* HM2* [Fs III]	99.61%	253	1	0	7.62E-133
Otu010	*F. intestinalis* JG1	99.61%	253	1	0	7.62E-133
Otu011	F. succinogenes HM2	92.49%	253	19	0	7.95E-103
Otu012	UW H9	97.23%	253	7	0	7.73E-123
Otu016	UW P2	96.46%	254	9	0	4.67E-120

Common OTUs were defined as those that were observed in more than one sample along with having a percent relative abundance of at least 0.1% in at least one sample.
Fibrobacter phylotype designations, if known, as well as type strain status (*) are indicated for references corresponding to OTUs containing less than, or equal to, one mismatch over the length of the sequence examined.

intestinalis group, as it exhibited close sequence similarity to *F. intestinalis* strain JG1 (Table 2.2). *F. intestinalis* strain JG1 was not included in our reference data set for classification to *Fibrobacter* phylotypes because of several ambiguous base calls in the full-length 16S rRNA gene sequences, however it was included in the maximum likelihood phylogeny inferred from near-full length 16S rRNA gene sequences (Figure 2.1). Otu003 evaded isolation, despite representing one of the most commonly observed Fibrobacteres OTUs in horse fecal samples. Otu003 was observed in 14 of 22 (63.64%) horse fecal samples, but also at a relatively low abundance in four of the seven (57.14%) Holstein cow rumen samples. Lastly, Fibrobacteres Otu004, likely representing the most phylogenetically divergent uncultured Fibrobacteres population observed in this study, was observed in four fecal samples from Holstein cows and in one fecal sample from a horse.

Phenotypic characteristics of *Fibrobacter* isolates

All of the 45 *Fibrobacter* strains exhibited growth on crystalline cellulose as the sole carbon source in the media formulations used (Table 2.3). However, strains UWP2 and UWR4 consistently produced lower concentrations of fermentation products and quantities of genomic DNA after extraction. In fact, UWR4 grew so poorly after isolation that fermentation products could not be reliably quantified for this strain. Attachment of at least a fraction of the population to cellulose during growth was observed via light microscopy for all strains. Generally, cultures reached cell densities between 1×10^8 - 1×10^{10} viable cells per mL when grown on cellulose based on the results of dilution to extinction in fresh media. No growth of any of the 45 strains was observed when xylan (from beechwood) was used as the sole carbon source in the growth medium. All strains produced succinate as the major fermentation product (5.78 ± 1.20 mM), (mean ± s.d.), with lesser amounts of acetate (3.42 ± 0.80 mM), and in some cases small amounts of formate (0.70 ± 0.42 mM). No lactate, propionate or butyrate production was detected (< 0.05 mM) for any of the 45 strains.

2.4 Discussion

Although the first isolates of *Fibrobacter* were reported over 50 years ago (33), cultured representatives for much of the apparent phylogenetic diversity of this group are lacking (20, 26, 27). As a result, knowledge regarding their ecology and physiology is limited. One barrier to achieving a more representative *Fibrobacter* culture collection is the technical challenge of the traditional anaerobic culture techniques used in the past to successfully recover isolates (34–36). Another is their preference for growth on an insoluble substrate, and requirement for attachment (19), which makes using agar-based media problematic. For this study, we developed a novel

Table 2.3: Media formulations used for isolation and characterization

Component	Isolation medium[#]	Characterization medium[$]
KH_2PO_4	10 mM	25 mM
Na_2CO_3	40 mM	-
$NaHCO_3$	-	40 mM
NaCl	15 mM	40 mM
L-Cysteine-HCl	6.35 mM	6.35 mM
NH_4Cl	15 mM	15 mM
Isobutyric acid	740 uM	740 uM
Isovaleric acid	640 uM	640 uM
Valeric acid	640 uM	640 uM
2-methylbutyric acid	650 uM	650 uM
$CaCl_2$	0.4 mM	0.4 mM
$MgCl_2$	0.5 mM	0.5 mM
$MnCl_2$	30 uM	30 uM
$FeSO_4$	50 uM	50 uM
$ZnCl_2$	10 uM	10 uM
$CoCl_2$	4 uM	4 uM
Na_2MoO_4	1 uM	1 uM
Na_2SeO_3	1 uM	1 uM
$NiCl_2$	5 uM	5 uM
Na_2WO_4	1 uM	1 uM
Nicotinamide	1.64 uM	1.64 uM
Pyridoxine HCl	973 nM	973 nM
Thiamine HCl	593 nM	593 nM
Riboflavin	531 nM	531 nM
Ca-D-pantothenate	420 nM	420 nM
p-Aminobenzoic acid	72.9 nM	72.9 nM
Biotin	20.5 nM	20.5 nM
Folic acid	2.83 nM	2.83 nM
Tetrahydrofolic acid	2.81 nM	2.81 nM
Vitamin B12	1.48 nM	1.48 nM
yeast extract	1 g/L	1 g/L
tryptone	1 g/L	1 g/L

[#]gas phase = 100% CO_2
[$]gas phase = 5% H_2, 20% CO_2, balance N_2

methodology for isolating these bacteria in order to alleviate some of these difficulties. Our approach leverages the tendency of *Fibrobacter* to adhere to cellulose (37, 38), which can be used for enrichment, followed by dilution to extinction, which has demonstrated utility for the isolation of bacteria that do not efficiently form colonies on agar plates (39). Using this method,

we successfully purified 45 axenic cultures of *Fibrobacter*, including the first isolated strains conclusively representative of four distinct phylotypes (Fs V, Fs VI, Fs VII, and Fi III). Our method proved very reliable for isolating relatively abundant *F. succinogenes* phylotypes from Holstein cow rumen samples, as isolates of either phylotype Fs I or Fs II were recovered from all 14 cow rumen samples subjected to isolation. Isolates of phylotype Fs III and Fs IV, typically present but at lower abundance in the rumen (20, 25), were not recovered from cow rumen samples, suggesting a potential bias towards the most numerically abundant strains in a given sample. Although representatives of Fs IV were not recovered from rumen samples, they were isolated from rhinoceros, tapir, and colobus monkey fecal samples, suggesting a broad host range for this phylotype. Isolates representing two phylotypes commonly observed in fecal samples from large hindgut-fermenting mammals, Fs V and Fs VI, were recovered from horse, rhinoceros, and elephant fecal samples. These isolates represent the first reported cultures of *Fibrobacter* from any rhinoceros and elephant hosts. Isolation of bacteria that fit the description of *Fibrobacter* have been previously reported for horses (40), and molecular evidence has suggested phylogenetically distinct populations in the hindgut of these animals (22). However, the horse isolates reported here provide the first conclusive synthesis of these previous observations. Representatives of *F. intestinalis* were also successfully isolated using our approach, with three phylotypes of *F. intestinalis* recovered from pigs, rhesus monkeys, and an ostrich. Strains of *F. intestinalis* have been previously isolated from pigs (36), but UW OS from the ostrich is the first reported isolate of *Fibrobacter* from a non-mammal.

Culture-independent analysis of the total bacterial community was also performed in order to investigate whether the *Fibrobacter* isolates we recovered accurately represent the dominant populations in their respective samples. In most instances, this was the case, especially

for the most extensively sampled hosts: horses and Holstein cows. Our culture-independent analysis did, however, provide strong evidence for *Fibrobacter* phylotypes in horse fecal and Holstein cow fecal samples that were not represented by any of our 45 isolates. A possible explanation for our inability to isolate these phylotypes is that the media formulation, which is based on a composition originally optimized for culturing rumen *Fibrobacter* strains (41), lacked specific growth factors. In particular, sterilized rumen fluid, or cecal extract, which have previously been used to aid in *Fibrobacter* isolations (34, 35, 40), was not included in our media in order to facilitate reproducibility by researchers who may lack access to these additives. As a result, we cannot exclude the possibility of missing, or suboptimal concentrations of, growth factors essential for certain strains or phylotypes. Most samples for which we were unable to obtain isolates did not display any signs of cells resembling *Fibrobacter* colonizing cellulose in primary enrichment cultures. Insufficient nutrition to stimulate and/or sustain growth would explain these observations, and might also explain why strains UWR4 and UWP2 consistently exhibited poor growth under the conditions used in this study despite their axenic status. Lastly, consideration must also be given to the possibility that at least some of these uncultured phylotypes are unable to grow efficiently on crystalline cellulose, which could also explains why axenic cultures were not recovered.

The combination of culture-dependent and culture-independent analyses used in this study provided new insights into the ecology of *Fibrobacter* spp. in the gastrointestinal tracts of herbivores. As expected, the highest relative abundances of Fibrobacteres sequences were observed in samples from strictly herbivorous hosts with large body weights. Importantly, the *Fibrobacter* populations in these animals were typically dominated by phylotypes from the *F. succinogenes* species group, and not *F. intestinalis*, despite most of these hosts being hindgut

fermenters. In fact, no sequences classified to any of the *F. intestinalis* phylotypes were observed in our 22 horse fecal samples. Although the highest abundance Fibrobacteres populations were typically composed of *F. succinogenes*-related phylotypes, an exception was the ostrich, which had a population almost entirely made up of phylotype Fi I. Lower relative abundances of *Fibrobacter* were observed in samples from pigs and several primates, with phylotype Fi I being the most commonly observed phylotype in these phylogenetically diverse hosts. It is not known why phylotype Fi I, which is represented by the *F. intestinalis* type strain NR9, is particularly suited for the hindguts of these hosts. No sequences classified as phylotype Fi I were observed in rumen samples examined here, or in those from the GRC. It has been suggested that, despite its name, *F. intestinalis* is also present in the rumen (20). Sequences likely representative of *F. intestinalis* strain JGI, originally isolated from the rumen of sheep (20), were observed in two sheep rumen contents examined here, as were sequences classified to phylotype Fi II in the GRC dataset, particularly in the rumen samples from giraffes. These observations suggest that some strains of *F. intestinalis* live in the rumen, but that they are likely distinct from those closely related to the *F. intestinalis* type strain and that their presence may depend on an association with certain host species.

A clear difference in the *Fibrobacter* phylotypes commonly observed in the bovine rumen and horse hindgut was observed in our 16S rRNA gene amplicon data. *Fibrobacter* phylotypes typically observed in the rumen, Fs I-IV, were either not detected, or rarely observed in horse fecal samples. Recovery and phylotyping of *Fibrobacter* isolates from horse fecal samples confirmed previously reported culture-independent evidence of phylogenetically distinct *Fibrobacter* populations in the hindgut of horses (22). We observed no evidence for populations of *Fibrobacter* phylotypes Fs V or Fs VI in rumen samples, although our culture-independent

data indicated that these phylotypes make up a substantial fraction of the Fibrobacteres population in the horse hindgut. Additionally, the total absence of sequences corresponding to these phylotypes in the GRC (32), a global dataset that includes hundreds of samples from diverse ruminant host species consuming different diets, across a broad geographic range, strongly reinforces that this is a general phenomenon. These phylogenetic and ecological attributes, separating them from previously characterized strains of *F. succinogenes*, suggest that phylotypes Fs V and Fs VI warrant reclassification as a distinct species of the genus *Fibrobacter*.

An underappreciated aspect of *Fibrobacter* ecology is the co-occurrence of multiple *Fibrobacter* phylotypes within a given host, as was demonstrated here for rumen samples as well as most fecal samples from horses. In Holstein cow rumen samples, sequences classified as *Fibrobacter* phylotypes Fs I, Fs II, Fs III, and Fs IV were consistently observed together within a given sample, although Fs I and Fs II were typically dominant. This pattern of *Fibrobacter* phylotype co-occurrence was found across multiple ruminant host species in the GRC dataset, and is in agreement with previous observations (20, 25, 42). We hypothesize that these *Fibrobacter* phylotypes occupy distinct ecological niches in vivo, but data supporting this supposition is sparse. Despite considerable phylogenetic diversity within the genus *Fibrobacter*, a general lack of phenotypic variation has been reported (20). All of the isolates examined in our study grew on crystalline cellulose, produced succinate and acetate as major fermentation products in similar ratios, and failed to exhibit any growth on beechwood xylan. Although it has been reported that some *Fibrobacter* strains are capable of growth on xylan (43), recent genomic evidence suggests that the inability of grow on pentoses is a unifying feature for the entire phylum (44). The expansion of cultured *Fibrobacter* representatives resulting from this work

will aid future efforts to identify the specific niches of distinct and co-occurring *Fibrobacter* phylotypes in the herbivore gastrointestinal tract.

In conclusion, this study not only expands our phylogenetic and physiological understanding of bacteria in the genus *Fibrobacter*, but also presents a novel and generalizable isolation method for recovering axenic cultures of these bacteria from herbivore hosts. Our study focused on hindgut-fermenting herbivores, as these hosts represent a rich source of uncultured *Fibrobacter* diversity and harbor microbial communities that are distinct and less well characterized than those of ruminants. Importantly, we present the first cultured representatives for four distinct *Fibrobacter* phylotypes, and our culture-independent analyses revealed insights into their ecology. Future work, including whole-genome sequencing and further phenotypic characterizations, promises to shed additional light on the ecology and evolution of these relatively abundant, but enigmatic, fiber-degrading bacterial symbionts of herbivores. Given the presence of Fibrobacteres within the gastrointestinal tracts of a broad diversity of herbivores, our work not only expands our understanding of this phylum, but also provides access to cultured isolates from which hypotheses generated from culture-independent, sequence-based studies can be effectively tested.

2.5 Materials & Methods

Sample collection and processing

Animal hosts sampled for this study were housed at the Henry Vilas Zoo (Madison, WI), the Milwaukee County Zoo (Milwaukee, WI), the US Dairy Forage Research Center farm (Prairie du Sac, WI), and several other research facilities managed by the University of Wisconsin – Madison (Madison, WI). Institutional Animal Care and Use Committee (IACUC)

approval was obtained from the proper authority where applicable. Fresh feces, or gastrointestinal contents where available, were collected aseptically in sterile 50 mL conical tubes, immediately placed on wet ice, and transported back to the lab for same day processing, typically less than four hours post collection. Upon arrival at the lab, samples were transferred to anaerobic conditions (atmosphere: 5% H_2, 20% CO_2, balance N_2) and preserved for future isolation attempts and culture-independent bacterial community analysis. Preservation of samples for culturing consisted of preparing a suspension by mixing one part sample with four parts sterile anaerobic buffer (155 mM NaCl, 3 mM Na_2HPO_4-$7H_2O$, 1.5 mM KH_2PO_4, 25 mM $NaHCO_3$, 6 mM L-cysteine HCl, pH=6.8) followed by 2x dilution in an equal volume of sterile glycerol, distribution to sterile serum vials with butyl rubber stoppers and aluminum crimp seals via needle and syringe, and storage at -80° C. Preservation of samples for culture-independent analysis involved 10x dilution in $T_{50}E_{50}$ sucrose buffer (50 mM Tris-HCl, 50 mM EDTA, 15% sucrose, pH = 8.0) and storage at -20° C.

Isolation of *Fibrobacter* spp.

Enrichment for anaerobic bacteria capable of growth on crystalline cellulose was achieved by serial dilution of the glycerol stocks prepared from feces, or gastrointestinal contents, in a slightly modified version of a medium originally described by Scott and Dehority (41) with 0.5% (wt vol^{-1}) Sigmacell 50 (Sigma-Aldrich, St. Louis, MO) as the primary carbon source. The exact formula, minus the carbon source, of the modified Dehority medium (MDM) used for enrichment and isolation can be found in Table 2.3. The final volume of media used was 10 mL in 18 x 150 mm anaerobic tubes, with a gas phase of 100% CO_2, sealed with gas impermeable butyl rubber stoppers and 20 mm aluminum crimp seals (Bellco Glass, Vineland,

NJ). The enrichments were incubated at 39° C and monitored via light microscopy daily for bacterial colonization of the cellulose crystals. Simple staining of the cells using crystal violet was employed to aid in microscopic examinations. Upon observing dense colonization, 1 mL of the enrichment was used to further enrich for cellulose-adherent bacteria using a modified version of the assay of adhesion described by Gong and Forsberg (37) under anaerobic conditions with a gas phase of 5% H_2, 20% CO_2, balance N_2. Briefly, 1 mL of sterile anaerobic buffer containing 2% (wt vol^{-1}) Sigmacell 50 (Sigma-Aldrich) was added to a 1.5 mL microcentrifuge tube and the cellulose was allowed to settle on the bottom of the tube. The liquid phase was discarded and 1 mL of enrichment culture was added to the cellulose pellet. The sample was mixed and bacterial attachment was promoted by incubating the tube at 39° C for 30 min. After the initial incubation period, the liquid phase was discarded and the cellulose pellet was washed with 1 mL of sterile anaerobic buffer (prewarmed to 39° C). The cellulose was allowed to settle, three mins, while maintaining 39° C. Once again, the liquid phase was discarded and the process was repeated for a total of ten washes to remove non-adherent bacteria. Upon completion of the washes, the sample was serially diluted in MDM + cellulose and incubated at 39° C. The enrichments were monitored daily for signs of growth characteristic of *Fibrobacter* cultures, including visual observation of a reduction in the size of the cellulose sediment, often accompanied by a surface layer of slime, and microscopic observation of dense colonization of the cellulose crystals by rod-shaped cells. The most dilute culture exhibiting the desired characteristics was selected and the process was repeated until a culture suspected to be axenic was obtained, typically 3-5 rounds of washes followed by dilution to extinction. Isolation was confirmed through the recovery of a single 16S rRNA gene sequence using the universal bacterial primers 27F and 1492R (29) (see below).

Phylogenetic analysis

Genomic DNA (gDNA) extraction from suspected axenic *Fibrobacter* cultures was performed using a combination of bead-beating and SDS for cell lysis followed by phenol:chloroform extraction and alcohol precipitation (45). Cell pellets from late log phase cultures, 24 – 48 h post inoculation, were resuspended in 1 mL DNA extraction buffer (100 mM Tris-HCl, 10 mM EDTA, 150 mM NaCl, pH = 8.0) and combined with 0.5 g of 0.1 mm zirconium beads, 700 µL equilibrated phenol (pH = 8.0) and 50 µL 20% SDS. The samples were subjected to bead-beating for 2 mins followed by incubation at 60° C for 10 mins followed by additional bead-beating for 2 min. The organic and aqueous phases were separated via centrifugation and the aqueous phase extracted with 500 µL equilibrated phenol:chloroform:isoamyl alcohol (25:24:1). A third, and final, extraction of the aqueous phase with 500 µL phenol:chloroform:isoamyl alcohol was performed followed by precipitation of the gDNA with 3 M Na acetate and isopropanol, then dried and subsequently resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH = 8.0). gDNA was quantified using the BR dsDNA assay kit and Qubit® Fluorometer (Invitrogen, Carlsbad, CA).

Near full-length 16S rRNA gene sequences were identified for each isolate by PCR and Sanger sequencing using the universal bacterial primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3') (29). PCR was performed using the Herculase II Fusion DNA Polymerase with dNTPs Combo kit (Agilent Technologies, Santa Clara, CA). Each 20 µL reaction contained: 4 µL 5x reaction buffer, 1 µL each of forward and reverse primers at a stock concentration of 10 µM, 0.2 µL 100 mM dNTPs, 0.4 µL Herculase II Fusion DNA Polymerase, 12.4 µL sterile H_2O, and 1 ng DNA

template. The cycling conditions were as follows: 1 min initial denaturation at 95° C followed by 30 cycles consisting of 30 seconds at 95° C, 45 seconds at 50° C, and 1 min at 72° C after which a final extension step for 10 mins at 72° C was performed. The PCR products, size ~1465 bp, were separated via electrophoresis through a 1% low-melt agarose (National Diagnostics, Atlanta, GA) gel and visualized using SYBR® Safe DNA Gel Stain (Invitrogen). PCR products of appropriate size were excised from the gel and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp., Irvine, CA). Purified PCR products were submitted for Sanger sequencing (Functional Biosciences, Inc., Madison, WI).

A maximum likelihood phylogenetic tree was inferred from near full-length 16S rRNA gene sequences in MEGA v7.0 (46) using the Tamura-Nei substitution model (47) and 1,000 bootstrap replicates. The initial tree for the heuristic search was generated automatically by applying Neighbor-Join and BioNJ algorithms to the matrix of pairwise distances estimated using maximum composite likelihood followed by selection of the phylogeny with the highest log likelihood. Prior to constructing the tree, the sequences were aligned in MEGA using ClustalW, with default parameters, and trimmed to an equal number of nucleotide positions (1,268 sites). A 16S rRNA gene sequence from the *Bacteroides fragilis* type strain NCTC 9343 (NCBI Ref. CR626927.1) was included as an outgroup. The operational taxonomic units (OTUs) represented by the *Fibrobacter* isolates, also referred to as *Fibrobacter* phylotypes, were determined using mothur v.1.38 (30). A distance matrix was created from the aligned and trimmed 16S rRNA gene sequences with the *dist.seqs* command and used as the input for the *cluster* command, which was run using the average neighbor algorithm with a distance cutoff of 0.03.

Culture-independent bacterial community profiling

Total DNA was extracted from gastrointestinal contents or feces that had been diluted and frozen in $T_{50}E_{50}$ sucrose buffer. The samples were thawed and the cells resuspended in 1 mL DNA extraction buffer. DNA was recovered using bead-beating and phenol:chloroform as described above for suspected isolates. PCR amplification of bacterial 16S rRNA DNA was performed using universal primers flanking the variable 4 (V4) region (31). Reactions contained 2 ng DNA template, 0.4 µM each primer, 12.5 µL 2x HotStart ReadyMix (KAPA Biosystems, Wilmington, MA), and H_2O to 25 µL. Cycling conditions were as follows: initial denaturation of 95° C for 3 mins followed by 25 cycles of 95° C for 30 seconds, 55° C for 30 seconds, and 72° C for 30 seconds, with a final extension at 72° C for 5 mins. The PCR products were separated via electrophoresis through a 1% low-melt agarose (National Diagnostics) gel and visualized using SYBR® Safe DNA Gel Stain (Invitrogen). PCR products of appropriate size were excised from the gel and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp.). Purified PCR products were quantified using the HS dsDNA assay kit and Qubit® Fluorometer (Invitrogen) and equimolar pooled. The pool plus 5% PhiX control DNA was sequenced with the MiSeq 2x250 v2 kit (Illumina, San Diego, CA) using custom sequencing primers (31).

Sequences were demultiplexed on the Illumina MiSeq and processed using mothur v.1.38 (30) in general accordance with the MiSeq S.O.P. (https://www.mothur.org/wiki/MiSeq_SOP). Individual contigs were assembled from the corresponding paired-end sequences and poor quality sequences were removed. Sequences were aligned to the SILVA 16S rRNA gene reference database (48) to check for alignment to the V4 target region. Chimeras were detected, and subsequently removed, using the chimera.uchime command

(http://drive5.com/usearch/manual/uchime_algo.html) on a trimmed version of the alignment that had been reduced using *unique.seqs* and *pre.cluster* (diffs = 2). Sequences that could not be classified to domain Bacteria, such as those classified to domain Eukaryota, domain Archaea, or domain "unknown" were removed from the dataset. Sequences suspected to be either host-cell or diet derived due to their classification as mitochondria and chloroplast, respectively, were also removed.

After the processing steps outlined above, the "cleaned" sequences were clustered into OTUs using the average neighbor algorithm with a distance cutoff of 0.03 with the *dist.seqs* and *cluster.split* commands in mothur. All OTUs represented by only a single sequence were subsequently removed using the *remove.rare* command. OTUs were classified using the Greengenes database (49) with the *classify.seqs* command (consensus confidence level ≥ 80%) followed by *classify.otu*. A matrix of OTU counts by sample, normalized to 10,000 sequences per sample, was generated using the *make.shared* and *normalize.shared* commands. A representative sequence for each OTU was extracted using the *get.oturep* command. The count data and relevant metadata were then imported into R (version 3.3.1) for statistical analyses and plotting (https://www.r-project.org/). A non-metric multidimensional scaling (nMDS) ordination analysis of Bray-Curtis dissimilarities among the total bacterial communities was performed using the *metaMDS* function from the "vegan" package (50). Standard error ellipses were generated for select groups of samples from the ordination results using the "vegan" function *ordiellipse*.

In silico analysis of *Fibrobacter* spp. ecology

V4 16S rRNA gene sequences classified to the phylum Fibrobacteres using Greengenes (49) were extracted from the "cleaned" fasta file using the *get.lineage* command in mothur. The Fibrobacteres sequences were then clustered into OTUs using the same methods described for the total bacterial data. The Fibrobacteres OTUs were then classified in mothur, as described above, to *Fibrobacter* phylotypes using the 16S rRNA gene sequences for the isolates described in this study, as well as *F. succinogenes* subsp. *elongatus* strain HM2 (NCBI Ref. NR_104844.1), and a taxonomy file which included their designated phylotypes. No other *Fibrobacter* sequences from previously reported isolates were included in the reference taxonomy due to either a lack of sequences of sufficient quality or because they failed to add phylotype diversity to the dataset. As a result, only sequences containing no ambiguous base calls from cultured isolates were included. A matrix of the Fibrobacteres OTU counts by sample was generated using the *make.shared* command. The Fibrobacteres OTU counts were then summed according to phylotype and imported into R for statistical analysis and plotting. Normalized sequence counts for the *Fibrobacter* phylotypes were fit onto the total bacteria NMDS plot, described above, using the *envfit* function from "vegan". Vectors corresponding to phylotypes exhibiting a fit with $P < 0.20$ were plotted onto the ordination. However, only vectors exhibiting a fit with $P < 0.05$ after adjustment using the Bonferroni correction were deemed statistically significant. Representative sequences for Fibrobacteres OTUs were extracted in mothur with the *get.oturep* command, followed by the creation of a distance matrix from the representative sequences using the *dist.seqs* command. The representative sequences and distance matrix were then imported into R using the *read.phyDat* function from the "phangorn" (51) package and the *import_mothur_dist* function from the "phyloseq" (52) package. A maximum likelihood neighbor-joining tree was inferred from the representative

sequences and distance matrix in R using the *bionj* function from the "ape" (53) package followed by the *pml*, *optim.pml*, and *bootstrap.pml* functions from the "phangorn" package. The phylogenetic tree and Fibrobacteres OTU count table were then used to create a phyloseq object using *phyloseq* from the "phyloseq" package, and the phylogenetic tree plotted with the function *plot_tree*. Closest BLAST hits for representative sequences for Fibrobacteres OTUs were determined by running *blastn* from the command line (54) using the near-full length 16S rRNA gene sequences of *Fibrobacter* isolates from this study and previously described reference strains (20) as the reference database.

16S rRNA gene amplicon sequence data from the Global Rumen Census (GRC) (32) was downloaded from the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA), https://www.ncbi.nlm.nih.gov/sra (PRJNA272135, PRJNA272136, PRJNA273417), using *fastq-dump* from the SRA toolkit. Raw sequences from the GRC were processed in mothur using a pipeline similar to the one described for the Illumina data generated in this study. Sequences classified to the phylum Fibrobacteres using Greengenes were extracted from the "cleaned" GRC fasta file using the *get.lineage* command. The Fibrobacteres sequences were then clustered into OTUs and classified in mothur to *Fibrobacter* phylotypes using the *Fibrobacter* custom taxonomy file, described above. The matrix of GRC Fibrobacteres OTU counts by sample was generated using the *make.shared* command. The GRC Fibrobacteres OTU counts were then summed according to phylotype and imported into R for statistical analysis and plotting.

Phenotypic characterization

Growth on crystalline cellulose and xylan was investigated using MDM with either 0.3% (wt vol^{-1}) Sigmacell 20 (Sigma-Aldrich, St. Louis, MO) or 0.3% (wt vol^{-1}) xylan from beechwood (Megazyme, Bray, Ireland) as the primary carbon source. All growth tests were performed in triplicate, with 1 mL of medium in 1.5 mL microcentrifuge tubes under anaerobic conditions with a gas phase of 5% H_2, 20% CO_2, balance N_2. The cultures were incubated at 39° C for 48 h. Growth was assessed by visually monitoring the cultures for increased turbidity. Fermentation products present in the supernatants of each triplicate 48 h culture for strains exhibiting growth on a particular carbon source were quantified using high-performance liquid chromatography as described previously (55). Concentrations of fermentation products for positive cultures are summarized as the average ± one standard deviation.

Accession numbers

Near full-length 16S rRNA gene sequences for the 45 *Fibrobacter* strains isolated in this study have been deposited in the NCBI Nucleotide database under accessions numbers KY463324 – KY463368. Raw Illumina sequencing reads corresponding to the V4 16S rRNA gene amplicons for the 83 animal host gut and feces microbiotas have been deposited in the National Center for Biotechnology Information's Short Read Archive database under Bioproject accession PRJNA362214.

2.6 References

1. Mackie RI. 2002. Mutualistic fermentative digestion in the gastrointestinal tract: diversity and evolution. Integr Comp Biol 42:319–326.
2. Morrison M, Pope PB, Denman SE, McSweeney CS. 2009. Plant biomass degradation by gut microbiomes: more of the same or something new? Curr Opin Biotechnol 20:358–363.
3. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol 6:121–131.
4. Leschine SB. 1995. Cellulose degradation in anaerobic environments. Annu Rev Microbiol 49:399–426.
5. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577.
6. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 6:776–788.
7. Flint HJ. 1997. The rumen microbial ecosystem--some recent developments. Trends Microbiol 5:483–488.
8. Cosgrove DJ. 2005. Growth of the plant cell wall. Nat Rev Mol Cell Biol 6:850–861.
9. Béguin P. 1990. Molecular biology of cellulose degradation. Annu Rev Microbiol 44:219–248.
10. Dehority BA, Scott HW. 1967. Extent of cellulose and hemicellulose digestion in various forages by pure cultures of rumen bacteria. J Dairy Sci 50:1136–1141.
11. Hungate RE. 1975. The rumen microbial ecosystem. Annual Review of Ecology and Systematics 6:39–66.
12. Kobayashi Y, Shinkai T, Koike S. 2008. Ecological and physiological characterization shows that *Fibrobacter succinogenes* is important in rumen fiber digestion - review. Folia Microbiol (Praha) 53:195–200.
13. Hungate RE. 1950. The anaerobic mesophilic cellulolytic bacteria. Bacteriol Rev 14:1–49.
14. Montgomery L, Flesher B, Stahl D. 1988. Transfer of *Bacteroides succinogenes* (Hungate) to *Fibrobacter* gen. nov. as *Fibrobacter succinogenes* comb. nov. and description of *Fibrobacter intestinalis* sp. nov. International Journal of Systematic and Evolutionary Microbiology 38:430–435.
15. Ransom-Jones E, Jones DL, McCarthy AJ, McDonald JE. 2012. The Fibrobacteres: an important phylum of cellulose-degrading bacteria. Microb Ecol 63:267–281.
16. Weimer PJ. 1996. Why don’t ruminal bacteria digest cellulose faster? J Dairy Sci 79:1496–1502.
17. Qi M, Jun H-S, Forsberg CW. 2007. Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. Appl Environ Microbiol 73:6098–6105.
18. Nouaille R, Matulova M, Pätoprstý V, Delort A-M, Forano E. 2009. Production of oligosaccharides and cellobionic acid by *Fibrobacter succinogenes* S85 growing on sugars, cellulose and wheat straw. Appl Microbiol Biotechnol 83:425–433.
19. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D, Brumm PJ.

2011. The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. PLoS ONE 6:e18814.

20. Amann RI, Lin C, Key R, Montgomery L, Stahl DA. 1992. Diversity among *Fibrobacter* isolates: towards a phylogenetic classification. Systematic and Applied Microbiology 15:23–31.
21. Stewart CS, Flint HJ. 1989. *Bacteroides* (*Fibrobacter*) *succinogenes*, a cellulolytic anaerobic bacterium from the gastrointestinal tract. Appl Microbiol Biotechnol 30:433–439.
22. Lin C, Stahl DA. 1995. Taxon-specific probes for the cellulolytic genus *Fibrobacter* reveal abundant and novel equine-associated populations. Appl Environ Microbiol 61:1348–1351.
23. Ilmberger N, Güllert S, Dannenberg J, Rabausch U, Torres J, Wemheuer B, Alawi M, Poehlein A, Chow J, Turaev D, Rattei T, Schmeisser C, Salomon J, Olsen PB, Daniel R, Grundhoff A, Borchert MS, Streit WR. 2014. A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. PLoS ONE 9:e106707.
24. Matsui H, Kato Y, Chikaraishi T, Moritani M, Ban-Tokuda T, Wakita M. 2010. Microbial diversity in ostrich ceca as revealed by 16S ribosomal RNA gene clone library and detection of novel *Fibrobacter* species. Anaerobe 16:83–93.
25. Shinkai T, Ohji R, Matsumoto N, Kobayashi Y. 2009. Fibrolytic capabilities of ruminal bacterium *Fibrobacter succinogenes* in relation to its phylogenetic grouping. FEMS Microbiol Lett 294:183–190.
26. Jewell KA, Scott JJ, Adams SM, Suen G. 2013. A phylogenetic analysis of the phylum Fibrobacteres. Syst Appl Microbiol 36:376–382.
27. Ransom-Jones E, Jones DL, Edwards A, McDonald JE. 2014. Distribution and diversity of members of the bacterial phylum Fibrobacteres in environments where cellulose degradation occurs. Syst Appl Microbiol 37:502–509.
28. Walker AW, Duncan SH, Louis P, Flint HJ. 2014. Phylogeny, culturing, and metagenomics of the human gut microbiota. Trends Microbiol 22:267–274.
29. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173:697–703.
30. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541.
31. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–5120.
32. Henderson G, Cox F, Ganesh S, Jonker A, Young W, Collaborators GRC, Abecia L, Angarita E, Aravena P, Arenas GN, Ariza C, Attwood GT, Avila JM, Avila-Stagno J, Bannink A, Barahona R, Batistotti M, Bertelsen MF, Brown-Kav A, Carvajal AM, Cersosimo L, Chaves AV, Church J, Clipson N, Cobos-Peralta MA, Cookson AL, Cravero S, Carballo OC, Crosley K, Cruz G, Cucchi MC, Barra R de la, Menezes ABD, Detmann E, Dieho K, Dijkstra J, Reis WLS dos, Dugan MER, Ebrahimi SH, Eythórsdóttir E, Fon FN, Fraga M, Franco F, Friedeman C, Fukuma N, Gagić D, Gangnat

I, Grilli DJ, Guan LL, Miri VH, Hernandez-Sanabria E, Gomez AXI, Isah OA, Ishaq S, Jami E, Jelincic J, Kantanen J, Kelly WJ, Kim S-H, Klieve A, Kobayashi Y, Koike S, Kopecny J, Kristensen TN, Krizsan SJ, LaChance H, Lachman M, Lamberson WR, Lambie S, Lassen J, Leahy SC, Lee S-S, Leiber F, Lewis E, Lin B, Lira R, Lund P, Macipe E, Mamuad LL, Mantovani HC, Marcoppido GA, Márquez C, Martin C, Martinez G, Martinez ME, Mayorga OL, McAllister TA, McSweeney C, Mestre L, Minnee E, Mitsumori M, Mizrahi I, Molina I, Muenger A, Muñoz C, Murovec B, Newbold J, Nsereko V, O'Donovan M, Okunade S, O'Neill B, Ospina S, Ouwerkerk D, Parra D, Pereira LGR, Pinares-Patiño C, Pope PB, Poulsen M, Rodehutscord M, Rodriguez T, Saito K, Sales F, Sauer C, Shingfield K, Shoji N, Simunek J, Stojanović-Radić Z, Stres B, Sun X, Swartz J, Tan ZL, Tapio I, Taxis TM, Tomkins N, Ungerfeld E, Valizadeh R, Adrichem P van, Hamme JV, Hoven WV, Waghorn G, Wallace RJ, Wang M, Waters SM, Keogh K, Witzig M, Wright A-DG, Yamano H, Yan T, Yáñez-Ruiz DR, Yeoman CJ, Zambrano R, Zeitz J, Zhou M, Zhou HW, Zou CX, Zunino P, 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.

33. Hungate RE. 1947. Studies on cellulose fermentation: III. The culture and isolation for cellulose-decomposing bacteria from the rumen of cattle. J Bacteriol 53:631–645.
34. Stewart CS, Paniagua C, Dinsdale D, Cheng KJ, Garrow SH. 1981. Selective isolation and characteristics of *Bacteriodes succinogenes* from the rumen of a cow. Appl Environ Microbiol 41:504–510.
35. Macy JM, Farrand JR, Montgomery L. 1982. Cellulolytic and non-cellulolytic bacteria in rat gastrointestinal tracts. Appl Environ Microbiol 44:1428–1434.
36. Varel VH, Fryda SJ, Robinson IM. 1984. Cellulolytic bacteria from pig large intestine. Appl Environ Microbiol 47:219–221.
37. Gong J, Forsberg CW. 1989. Factors affecting adhesion of *Fibrobacter succinogenes* subsp. *succinogenes* S85 and adherence-defective mutants to cellulose. Appl Environ Microbiol 55:3039–3044.
38. Minato H, Suto T. 1978. Technique for fractionation of bacteria in rumen microbial ecosystem. J Gen Appl Microbiol 24:1–16.
39. Kenters N, Henderson G, Jeyanathan J, Kittelmann S, Janssen PH. 2011. Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. J Microbiol Methods 84:52–60.
40. Davies ME. 1964. Cellulolytic bacteria isolated from the large intestine of the horse. Journal of Applied Bacteriology 27:373–378.
41. Scott HW, Dehority BA. 1965. Vitamin requirements of several cellulolytic rumen bacteria. J Bacteriol 89:1169–1175.
42. Shinkai T, Ueki T, Kobayashi Y. 2010. Detection and identification of rumen bacteria constituting a fibrolytic consortium dominated by *Fibrobacter succinogenes*. Animal Science Journal 81:72–79.
43. 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*. Can J Microbiol 39:780–786.
44. Abdul Rahman N, Parks DH, Vanwonterghem I, Morrison M, Tyson GW, Hugenholtz P. 2015. A phylogenomic analysis of the bacterial phylum Fibrobacteres. Front Microbiol 6:1469.

45. Stevenson DM, Weimer PJ. 2007. Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Appl Microbiol Biotechnol 75:165–174.
46. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874.
47. Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526.
48. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196.
49. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072.
50. Dixon P. 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science 14:927–930.
51. Schliep KP. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593.
52. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8:e61217.
53. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290.
54. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
55. Weimer PJ, Shi Y, Odt CL. 1991. A segmented gas/liquid delivery system for continuous culture of microorganisms on insoluble substrates and its use for growth of *Ruminococcus flavefaciens* on cellulose. Appl Microbiol Biotechnol 36:178–183.

Chapter 3: Carbohydrate-active enzyme gene content is correlated with the phylogeny of herbivore-associated *Fibrobacter* spp.

Authors and contributors:

Anthony Neumann: Designed the experiments, performed the experiments, analyzed the data, and wrote the manuscript

Dr. Garret Suen: Designed the experiments, and edited the manuscript

Funding:

This work was supported by a traineeship from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number T32AI55397 to APN, a UW-Madison Richard M. Heins Wisconsin Distinguished Graduate Fellowship to APN, a U.S. Department of Energy Joint Genome Institute Community Science Program sequencing grant to APN and GS, and a US Department of Energy Biological and Environmental Research Early Career Research Program Award DE-SC0008104 to GS. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The funders had no input in the design, execution, or interpretation of this work.

3.1 Abstract

Members of the genus *Fibrobacter* are cellulose-degrading bacteria and common constituents of the gastrointestinal microbiotas of herbivores. Although considerable phylogenetic diversity is observed among members of this group, few functional differences explaining the distinct ecological distributions of specific phylotypes have been described. In this study we performed a comparative analysis of whole-genome sequences generated from 38 novel *Fibrobacter* strains and the type strains for the two formally described *Fibrobacter* species: *F. succinogenes* strain S85 and *F. intestinalis* strain NR9. Significant differences in the number of genes encoding carbohydrate-active enzyme families involved in plant cell wall polysaccharide degradation were observed among *Fibrobacter* phylotypes. *F. succinogenes* genomes were consistently enriched in genes encoding carbohydrate-active enzymes compared to those of *F. intestinalis* strains. Moreover, genomes of *F. succinogenes* phylotypes that are dominant in the rumen had significantly more genes annotated to major families involved in hemicellulose degradation by *Fibrobacter* (*e.g.*, CE6, GH10, and GH43) than did the genomes of *F. succinogenes* phylotypes commonly observed in the lower gut of large hindgut-fermenting herbivores such as horses. Furthermore, genes encoding a putative urease were identified in the genomes of 12 *Fibrobacter* strains, which were primarily isolated from hindgut-fermenting hosts. Screening for growth on urea as the sole source of nitrogen provided strong evidence that the urease was active in these strains. These results represent a considerable contribution to our understanding of the ecology of *Fibrobacter* spp. in the herbivore gut.

3.2 Introduction

Herbivorous animals are capable of living on a diet of recalcitrant plant cell wall polysaccharides (PCWP) because of a symbiotic association with fibrolytic microbial communities residing in their gastrointestinal tract (1). Animals generally lack the ability to hydrolyze PCWP themselves, and are therefore dependent on their gut microbes to degrade and ferment these substrates into short-chain fatty acids, which they can then use for energy (2). Efficient digestion of PCWP requires a complex and dynamic microbiota because the plant cell wall has evolved a structure that is heterogeneous, chemically diverse, and highly resistant to depolymerization (3). The plant cell wall consists of cellulose fibers embedded in a matrix of hemicellulose, pectin, and lignin, and fibrolytic microbes produce a suite of enzymes that work synergistically to degrade these different classes of organic polymers (4). These carbohydrate-active enzymes (CAZymes) have been classified into numerous classes and families based on their function and homology (5, 6).

The major classes of catalytic CAZymes are: glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activities (AAs) (7). GHs hydrolyze glycosidic bonds in polysaccharides, in either an endo- or exo- manner. GTs catalyze the formation of glycosidic linkages from activated carbohydrate precursors. PLs perform non-hydrolytic cleavage of uronic acid-containing polysaccharides, such as those found in pectins. CEs hydrolyze carbohydrate esters, and include enzymes for debranching hemicellulose which is often highly substituted with acetyl and phenolic residues. AAs encompass enzymes involved in redox reactions that enhance the activity of other catalytic CAZymes. Catalytic CAZymes are often covalently linked to a non-catalytic domain that has an affinity for binding certain carbohydrates. These carbohydrate-

binding modules (CBMs) enhance the efficiency of their catalytic partners by ensuring close contact between the enzyme and its target substrate. Multiple enzymes from most, if not all, of these CAZyme classes are produced by the most actively fibrolytic microorganisms (8–11).

Bacteria possessing a superior ability to solubilize PCWP are often particularly well suited to utilize cellulose for growth (12, 13). Cellulose is commonly the most abundant structural polysaccharide in plant matter, but it is also among the most resistant to degradation (14). The relationship between herbivores and cellulose-utilizing bacteria is best understood in the context of the bovine rumen, where species of *Ruminococcus* and *Fibrobacter* are thought to be the most important primary degraders (15). *Fibrobacter* is a conserved member of the rumen microbiota, and these bacteria possess several characteristics that contribute to efficient rumen function (16). Although *Fibrobacter* readily solubilizes hemicellulose and pectin, likely as a means to gain access to cellulose, available evidence indicates that these bacteria only utilize hexoses resulting from the breakdown of cellulose for growth (17–19). As a result, xylooligosaccharides and pentoses liberated from insoluble forms of hemicellulose and pectin via *Fibrobacter* become available for consumption by other rumen microbes. *Fibrobacter* also contributes to rumen efficiency through the production of succinate, the primary fermentation product produced by these bacteria (20). Succinate does not accumulate in the rumen and is converted primarily to propionate, an important substrate used by the host for gluconeogenesis (21–23). This pathway preserves reducing equivalents for energy generation by the host that might otherwise be lost to methane production through the consumption of hydrogen, an alternative fermentation product, by methanogens (24).

Since first being isolated from the rumen, evidence for diverse populations of *Fibrobacter*, and related bacteria in the phylum Fibrobacteres, has accumulated as a result of

culture-independent studies (25–28). Still, potential functional differences explaining much of this diversity is lacking due, at least in part, to a general absence of axenic cultures available for study (29). Two species of *Fibrobacter*, *F. intestinalis* and *F. succinogenes*, have been formally described, but these species are not reliably differentiated using traditional characterization techniques on available representative strains (30). However, comparative genomics of Fibrobacteres genomes recovered from metagenomic surveys has recently demonstrated the utility of this approach for gaining functional insights into phylogenetic and ecological differences among members of this poorly understood phylum (31). We recently reported a novel isolation technique that was used to greatly expand the diversity of *Fibrobacter* strains available for study, as well as provide additional evidence strengthening the association of certain phylotypes with the rumen and others with the hindgut (17). The goal of this study was to build upon our previous work by making functional predictions explaining the ecological basis for diversity among *Fibrobacter* spp. in the herbivore gut using whole-genome comparisons of our previously described *Fibrobacter* isolates. We hypothesized that gastrointestinal location as well as differences in substrate preferences are responsible for different ecological distributions observed among phylotypes. Our results indicate a strong association between the *Fibrobacter* phylogeny and CAZyme gene content which likely contributes to the ecology of these bacteria, especially the dominance of certain phylotypes in the rumen.

3.3 Results

A total of 40 *Fibrobacter* genomes from axenic cultures were analyzed in this study, including type strains *F. succinogenes* S85 and *F. intestinalis* NR9 as well as 38 strains recently isolated from diverse herbivores (Table 3.1) (17). Although a draft genome sequence for

Table 3.1: *Fibrobacter* strains for comparative genomics

Strain	Phylotype	Host (ID)	Source	Sample site
UWB1	Fs II	Bovine (6844)	rumen	DairyCattleCenter (WI, USA)
UWB2	Fs I	Bovine (4270)	rumen	USDA PrairieDuSac (WI, USA)
UWB3	Fs I	Bovine (4291)	rumen	USDA PrairieDuSac (WI, USA)
UWB4	Fs I	Bovine (4413)	rumen	USDA PrairieDuSac (WI, USA)
UWB5	Fs II	Bovine (4413)	rumen	USDA PrairieDuSac (WI, USA)
UWB6	Fs II	Bovine (5871)	rumen	DairyCattleCenter (WI, USA)
UWB7	Fs I	Bovine	rumen	UW VeterinaryHospital (WI, USA)
UWB8	Fs II	Bovine (5971)	rumen	UW DairyCattleCenter (WI, USA)
UWB10	Fs II	Bovine	rumen	UW VeterinaryHospital (WI, USA)
UWB11	Fs I	Bovine (4280)	rumen	USDA PrairieDuSac (WI, USA)
UWB12	Fs I	Bovine (4296)	rumen	USDA PrairieDuSac (WI, USA)
UWB13	Fs I	Bovine (4297)	rumen	USDA PrairieDuSac (WI, USA)
UWB15	Fs II	Bovine (6107)	rumen	DairyCattleCenter (WI, USA)
UWB16	Fs I	Bovine (6551)	rumen	UW DairyCattleCenter (WI, USA)
UWCM	Fs IV	Colobus	feces	MilwaukeeCountyZoo (WI, USA)
UWEL	Fs VI	Elephant	feces	MilwaukeeCountyZoo (WI, USA)
UWH1	Fs V	Equine	feces	UW CharmanyInstructionalFacility (WI, USA)
UWH3	Fs V	Equine	feces	UW CharmanyInstructionalFacility (WI, USA)
UWH4	Fs II	Equine	feces	RollFarm (WI, USA)
UWH5	Fs V	Equine	feces	PurselyStables (WI, USA)
UWH6	Fs V	Equine	feces	UW CharmanyInstructionalFacility (WI, USA)
UWH8	Fs V	Equine	feces	UW CharmanyInstructionalFacility (WI, USA)
UWH9	Fs V	Equine	feces	UW VeterinaryHospital (WI, USA)
UWOS	Fi I	Ostrich	feces	HenryVilasZoo (WI, USA)
UWOV1	Fs II	Ovine	rumen	UW Meat&Muscle lab (WI, USA)
UWP2	Fs VII	Capybara	feces	HenryVilasZoo (WI, USA)
UWR1	Fs VI	Rhino	feces	HenryVilasZoo (WI, USA)
UWR2	Fs IV	Rhino	feces	HenryVilasZoo (WI, USA)
UWR3	Fs IV	Rhino	feces	MilwaukeeCountyZoo (WI, USA)
UWR4	Fs VI	Rhino	feces	HenryVilasZoo (WI, USA)
UWRM	Fi III	RhesusMacaque	feces	UW PrimateResearchCenter (WI, USA)
UWS1	Fi I	Porcine	feces	UW AnimalSciences (WI, USA)
UWS2	ND	Porcine	cecum	UW Meat&Muscle lab (WI, USA)
UWS3	Fi II	Porcine	cecum	UW Meat&Muscle lab (WI, USA)
UWS4	Fi III	Porcine	cecum	UW Meat&Muscle lab (WI, USA)
UWT1	Fs V	MalayanTapir	feces	HenryVilasZoo (WI, USA)
UWT2	Fs II	MalayanTapir	feces	HenryVilasZoo (WI, USA)
UWT3	Fs IV	Baird's	feces	MilwaukeeCountyZoo (WI, USA)
S85*	Fs I	Bovine	rumen	NA
NR9*	Fi I	Rat	cecum	NA

*Type strain
NA = not available

the *F. intestinalis* type strain, NR9, was publicly available prior to this work, this strain was re-sequenced with the aim of improving the assembly. This resulted in an improved draft of NR9, consisting of a single scaffold, which was subsequently utilized for all genomic analyses of NR9

Table 3.2: *Fibrobacter* genome statistics

Strain	Completeness	Genome size (Mb)	GC	# scaffolds	# predicted genes	Proteins
UWB1	98.9	3.601604	50.2	46	2925	2904
UWB2	100	3.528394	48.8	15	2907	2898
UWB3	100	3.570823	48.4	20	2892	2874
UWB4	100	3.586285	48.2	18	2915	2901
UWB5	100	3.361559	50.9	17	2717	2709
UWB6	100	3.46817	50.2	26	2823	2806
UWB7	100	3.812257	47.9	12	3110	3098
UWB8	100	3.46935	50.2	30	2827	2810
UWB10	100	3.479813	49.4	10	2785	2773
UWB11	100	3.763481	46.8	9	3063	3055
UWB12	100	3.826342	46.8	21	3096	3078
UWB13	100	3.804669	47.9	10	3103	3096
UWB15	100	3.464091	50.2	24	2821	2804
UWB16	100	3.555739	48.5	9	2884	2874
UWCM	99.94	3.784167	53.4	65	3122	3095
UWEL	100	3.530042	49.1	62	2938	2910
UWH1	100	3.830823	48.9	96	3249	3234
UWH3	100	3.868941	48.8	62	3224	3197
UWH4	98.9	3.696015	50.8	22	3083	3056
UWH5	100	3.736655	49.1	59	3116	3081
UWH6	100	3.921019	48.8	50	3273	3236
UWH8	100	3.851655	49	97	3285	3246
UWH9	100	3.784473	49	52	3174	3140
UWOS	99.94	3.242118	48	93	2879	2829
UWOV1	98.9	3.687051	50.3	29	3062	3054
UWP2	100	3.234975	53.4	72	2719	2690
UWR1	100	3.606749	48.7	100	3037	2988
UWR2	100	3.278051	53.4	20	2721	2711
UWR3	99.94	3.546738	53.8	20	2929	2918
UWR4	100	3.60147	48.7	93	3004	2981
UWRM	99.94	2.85944	44.8	69	2578	2521
UWS1	99.94	3.380004	48.1	2	2950	2934
UWS2	100	2.947569	48.9	77	2661	2635
UWS3	100	3.127022	49.2	1	2836	2828
UWS4	98.84	2.849881	44.8	78	2553	2502
UWT1	100	4.019388	49.3	1	3432	3422
UWT2	100	3.493574	50.2	49	2785	2785
UWT3	99.94	3.810443	53.8	5	3225	3217
S85*	100	3.842635	48.1	1	3126	3119
NR9*	99.39	3.390891	47.9	1	2992	2979

*Type strain

in this study (Table 3.2). Overall, 16 of the *Fibrobacter* genomes were from strains isolated from rumen samples, while the remaining 24 were from strains isolated from the feces or cecal contents of hindgut-fermenting herbivores. The 40 strains were previously identified as

phylogenetically related to *F. succinogenes* (n = 33) and *F. intestinalis* (n = 7) (17). The genomes are estimated to be ≥ 98.8% complete, with all the assemblies consisting of 100 or fewer scaffolds (Table 3.2). The sizes of the genomes range from 2.85 to 4.02 Mb (mean = 3.56 ± 0.28 Mb). The GC content of the genomes range from 44.8% to 53.8% (mean = 49.3% ± 2.08%).

Fibrobacter strains from herbivores are phylogenetically diverse

Average nucleotide identity (ANI) for all pairwise comparisons of the 40 *Fibrobacter* genomes was calculated as a measure of overall diversity among the strains. The mean and median ANI for all comparisons was 82.5% ± 4.7% and 80.7%, respectively. Strains UWR1 and UWR4 were the most closely related based on ANI (99.9%), while strains NR9 and UWH5 were the most divergent based on ANI (76.1%). The 40 *Fibrobacter* strains were clustered based on ANI distance and grouped using a cutoff of 95% ANI in order to estimate the total number of individual species present (32). A total of 21 different species were identified among the 40 *Fibrobacter* strains using these criteria (Table 3.3). The largest species group identified contained 6 strains isolated from fecal samples originating from multiple horses and farms, as well as 1 strain isolated from tapir feces. No other species group contained more than 3 strains. Ten of the species groups were represented by only a single strain.

A phylogenetic model was constructed from concatenated alignments of single copy essential genes (33) in order to more clearly resolve the evolutionary relationships among the strains (Figure 3.1). The concatenated alignment consisted of 99 protein sequences and 36,135 amino acid positions. The 40 *Fibrobacter* strains clearly segregated into two groups with one group containing the 7 *F. intestinalis* strains and the other containing the 33 *F. succinogenes*

Table 3.3: *Fibrobacter* strain characteristics

Strain	95% ANI group	D-glucose	D-cellobiose	D-lactose	urea
UWB1	3	+	+	ND	ND
UWB2	4	+	+	ND	ND
UWB3	5	+	+	+	ND
UWB4	6	+	+	+	ND
UWB5	7	+	+	ND	ND
UWB6	13	+	+	ND	ND
UWB7	12	+	+	+	+
UWB8	13	+	+	ND	ND
UWB10	21	+	+	ND	+
UWB11	11	+	+	+	ND
UWB12	11	+	+	ND	ND
UWB13	12	+	+	ND	ND
UWB15	13	+	+	ND	ND
UWB16	5	+	+	ND	ND
UWCM	14	+	+	ND	ND
UWEL	15	+	+	ND	+
UWH1	8	+	+	ND	+
UWH3	8	+	+	ND	+
UWH4	16	+	+	ND	ND
UWH5	8	+	+	ND	+
UWH6	8	+	+	ND	+
UWH8	8	+	+	ND	+
UWH9	8	+	+	+	+
UWOS	1	+	+	ND	ND
UWOV1	3	+	+	ND	ND
UWP2	17	+	+	ND	ND
UWR1	19	+	+	ND	+
UWR2	9	+	+	ND	ND
UWR3	14	+	+	ND	ND
UWR4	19	+	+	ND	ND
UWRM	20	+	+	ND	ND
UWS1	1	+	+	ND	ND
UWS2	10	+	+	ND	ND
UWS3	10	+	+	ND	ND
UWS4	20	+	+	ND	ND
UWT1	8	+	+	ND	+
UWT2	18	+	+	ND	+
UWT3	14	+	+	ND	ND
S85	2	+	+	+	ND
NR9	1	+	+	ND	ND

No strains grew on L-arabinose, D-mannose, D-xylose, D-galactose, D-maltose, or L-rhamnose
ANI = average nucleotide identity
+ = growth observed
ND = growth not detected

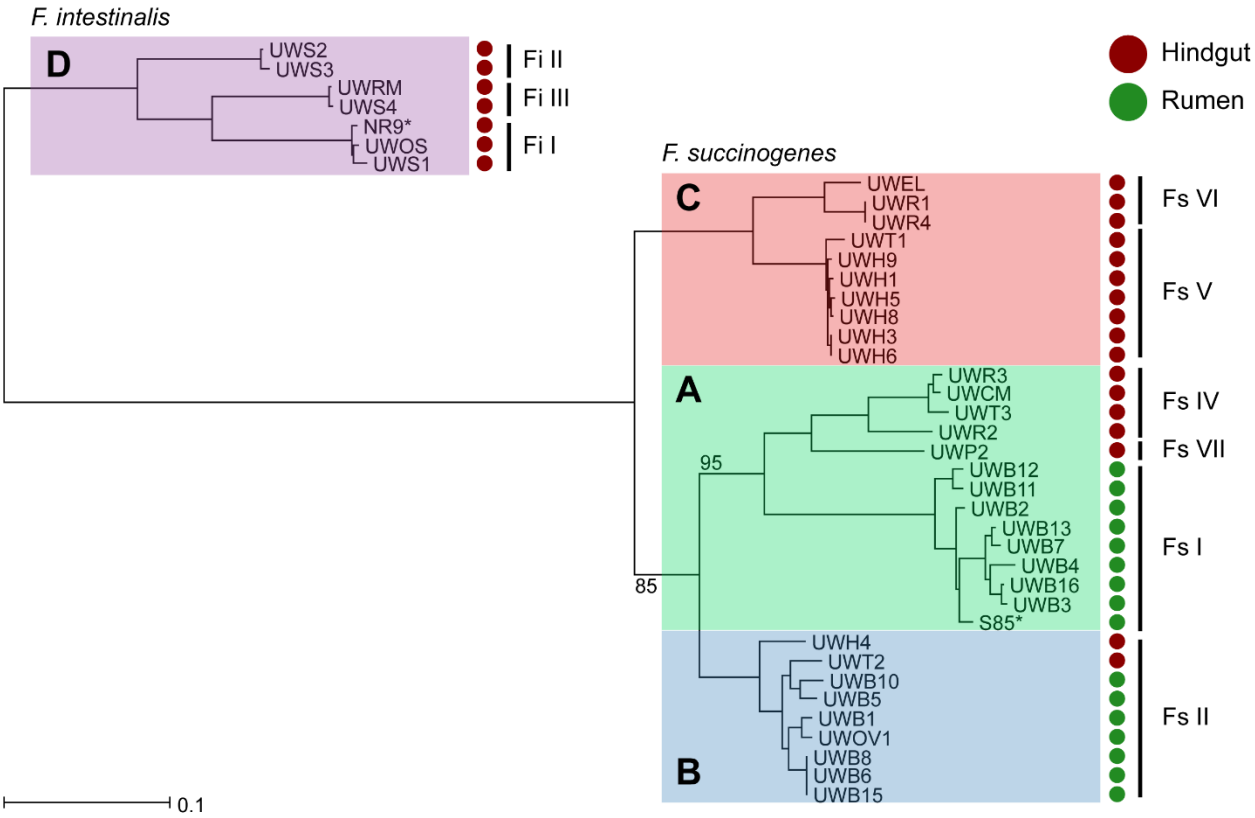


Figure 3.1: Essential gene phylogeny of *Fibrobacter* strains. The maximum likelihood phylogeny was constructed from the concatenated alignments of 99 essential protein genes using RAxML (69) and is midpoint rooted. Strains are annotated with circles according to their isolation source, rumen (dark green) or hindgut (dark red). Previously assigned *Fibrobacter* phylotypes are labeled and their coverage of the tree marked with black vertical bars (17). Type strains are identified with an asterisk (*). Major clades are shaded according to the following clade designations: clade A (green), clade B (blue), clade C (red), and clade D (violet). Bootstrap values for clades containing more than 5 strains are 100%, unless otherwise labeled (100 replicates). The scale represents the number of substitutions per site.

strains. The strains were further resolved into four readily distinguishable major clades, three of which contained all strains of *F. succinogenes* with the remaining clade containing all 7 strains of *F. intestinalis*. The four major clades were designated as *Fibrobacter* clades A - D (Figure 3.1). *Fibrobacter* clade A contained strains isolated from the rumen as well as from hindgut fermenters and consisted of the previously described phylotypes Fs I, Fs IV, and Fs VII (17). *Fibrobacter* clade B also contained rumen and hindgut isolates and was made up of the

previously described phylotype Fs II. The third *F. succinogenes* clade, clade C, was made up entirely of hindgut strains and phylotypes Fs V and Fs VI. Clade D consisted of all the *F. intestinalis* strains examined in this study, which were further resolved into three discrete groups coinciding with the previous designations of Fi I, Fi II, and Fi III.

A conserved general mechanism for PCWP degradation by *Fibrobacter* spp.

Predicted proteins in the *Fibrobacter* genomes were annotated to CAZyme families using the dbCAN database (34) in order to gain insights into the mechanisms of PCWP degradation utilized by the various strains. The total number of CAZymes predicted in the genomes ranged from a max of 219 for UWT2 and a min of 120 for UWOS. The average number of total CAZymes predicted in a single *Fibrobacter* genome was 190 ± 30.4, and the median was 198 (Table 3.4). Mean gene counts per genome for CAZyme classes predicted to play a role in PCWP degradation by *Fibrobacter* spp. were as follows: 85.1 ± 17.3 for glycoside hydrolases (GHs), 35.0 ± 7.6 for carbohydrate-binding modules (CBMs), 15.2 ± 4.1 for carbohydrate esterases (CEs), and 12.5 ± 2.6 for polysaccharide lyases (PL).

The specific CAZyme families identified in the genomes were generally consistent among the strains. Of the 72 CAZyme families identified among the 40 genomes, 39 (54.2%) were present in all *Fibrobacter* strains (Figure 3.2). These 39 conserved CAZyme families accounted for 87.6% of the total CAZymes identified. GH families containing the major cellulases produced by *Fibrobacter* (18), GH5, GH8, GH9, and GH45, were relatively abundant and conserved in all *Fibrobacter* genomes (Figure 3.3). A single copy of the gene predicted to encode the major cellulose-binding endoglucanase, consisting of CBM11 and GH51 domains, was identified in all 40 *Fibrobacter* genomes. Similar CAZyme families implicated in

Table 3.4: Total CAZyme summary statistics

Class	Gene count	Mean	Standard deviation	Median
CBM	1,398	35.0	7.6	36.5
CE	607	15.2	4.1	16.0
GH	3,404	85.1	17.3	90.0
GT	1,680	42.0	4.1	41.5
PL	5,01	12.5	2.6	13.0
AA	ND			
Total	7,590	190.0	30.4	198.0

ND = not detected

hemicellulose degradation in *Fibrobacter* (18) were present among the genomes (Figure 3.4). CBM family 6 and CBM family 35 were the most abundant carbohydrate-binding modules regardless of strain and were typically associated with the hemicellulose-degrading CAZyme families: CE6, GH10, GH26, and GH43 (Figure 3.5). Additional proteins thought to assist in cellulose degradation by *Fibrobacter* by facilitating cellular adherence are those containing fibro-slime domains (35). Genes predicted to encode fibro-slime domains were identified in all 40 *Fibrobacter* genomes. The fewest number of fibro-slime domain-containing proteins predicted in a single genome was 3, for several strains of *F. intestinalis*, while most strains of *F. succinogenes* had 8 - 10 distinct proteins predicted to contain these domains. A single copy of the gene for cellobiose/cellodextrin phosphorylase, an important enzyme for generating glucose-1-phosphate from cellodextrins resulting from the breakdown of cellulose was also identified in all 40 *Fibrobacter* genomes.

CAZyme genes are more abundant in *F. succinogenes* genomes

Clustering of the *Fibrobacter* strains according to the relative abundances of all CAZyme family genes in the genomes was performed in order to identify potential relationships among

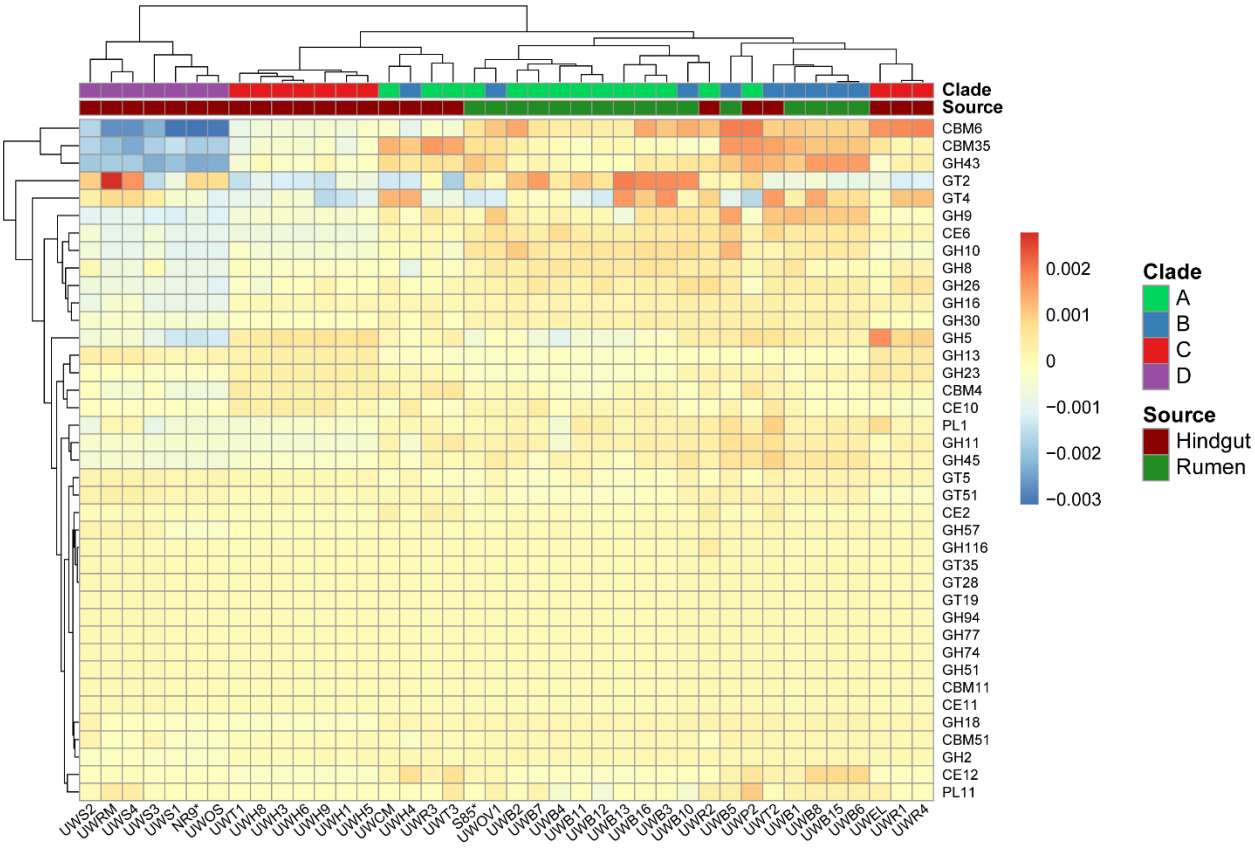


Figure 3.2: Variation among conserved CAZyme families in *Fibrobacter* genomes. Values are deviations from the mean relative abundance for a given CAZyme family. *Fibrobacter* strains are arranged horizontally according to the results of hierarchical clustering, according to conserved CAZyme genomic content, and labeled at the bottom of the plot. Strains are colored at the top of the plot according to phylogenetic clade membership: clade A (green), clade B (blue), clade C (red), and clade D (violet), and isolation source: rumen (dark green) or hindgut (dark red). CAZyme families are arranged vertically according to the results of hierarchical clustering with labels on the right side of the plot.

CAZyme content and the phylogeny. 2D nonmetric multidimensional scaling (NMDS) analysis identified a clear difference between the *F. intestinalis* strains (clade D) and the *F. succinogenes* strains (clades A - C) in CAZyme family gene content (Figure 3.6). Complete-linkage hierarchical clustering confirmed this result. Plots of total CAZyme genes and total genes annotated to each CAZyme class, normalized to control for differences in genome size, identified clear differences among the four major clades identified in the essential gene phylogeny (Figure 3.7). The largest differences were observed between the *F. succinogenes* clades

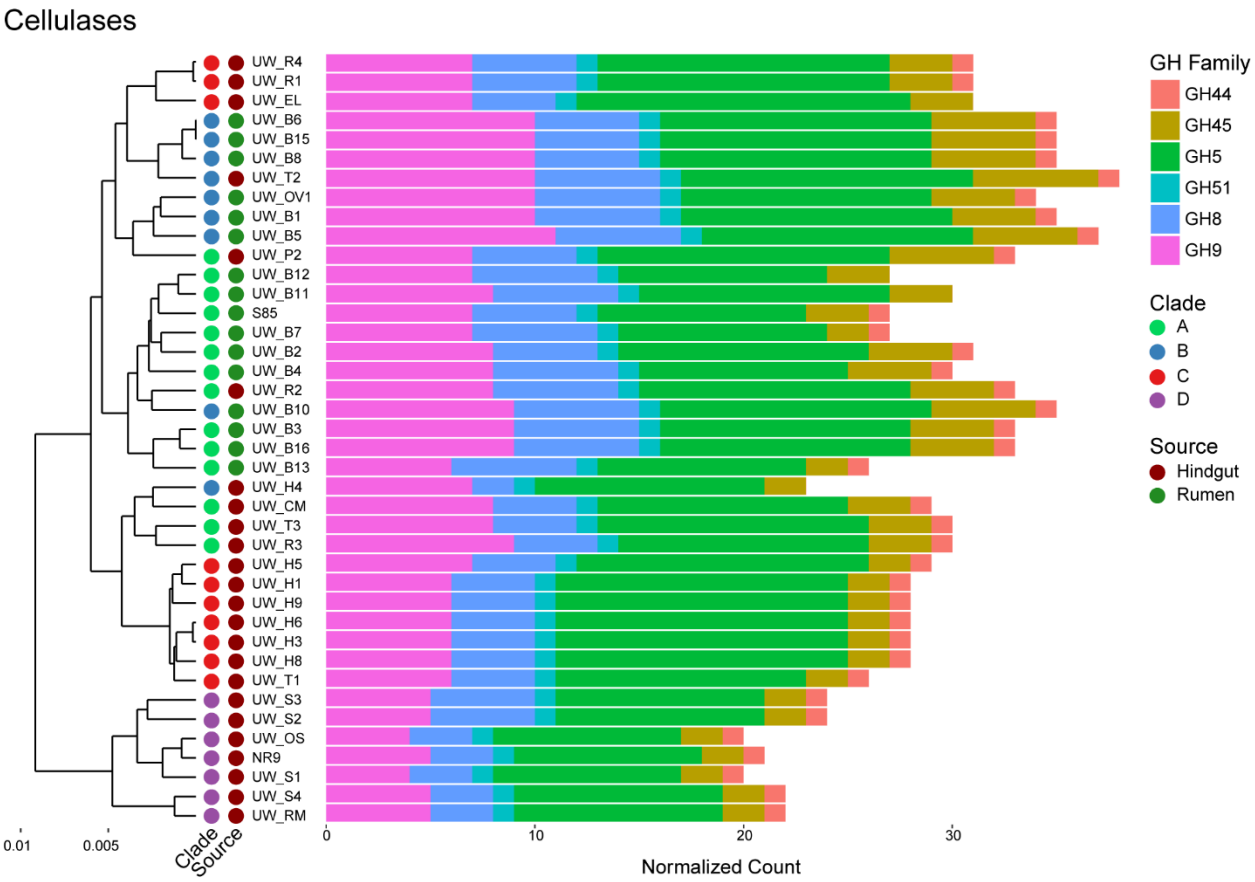


Figure 3.3: Distribution of cellulases among *Fibrobacter* strains. *Fibrobacter* strains are arranged according to the results of hierarchical clustering by relative abundance of all CAZyme family genes in their respective genomes. *Fibrobacter* strains are annotated with circles according to phylogenetic clade membership: clade A (green), clade B (blue), clade C (red), and clade D (violet), and isolation source: rumen (dark green) or hindgut (dark red). Horizontal bar plots represent the number of normalized gene counts in each glycoside hydrolase (GH) family.

(A - C) and the *F. intestinalis* clade (D). Significantly higher normalized gene counts were observed for each of the *F. succinogenes* clades A, B, and C compared to clade D for total CAZymes, CBMs, CEs, GHs, and PLs (Figure 3.7). A total of 20, out of 72, individual CAZyme families had significantly greater gene counts in each of clades A, B, and C when compared to clade D ($P < 0.05$). Among these were 13 of the 29 (44.8%) total GH families identified among the 40 *Fibrobacter* genomes. CBM family 77 and GH family 13 were the only CAZyme

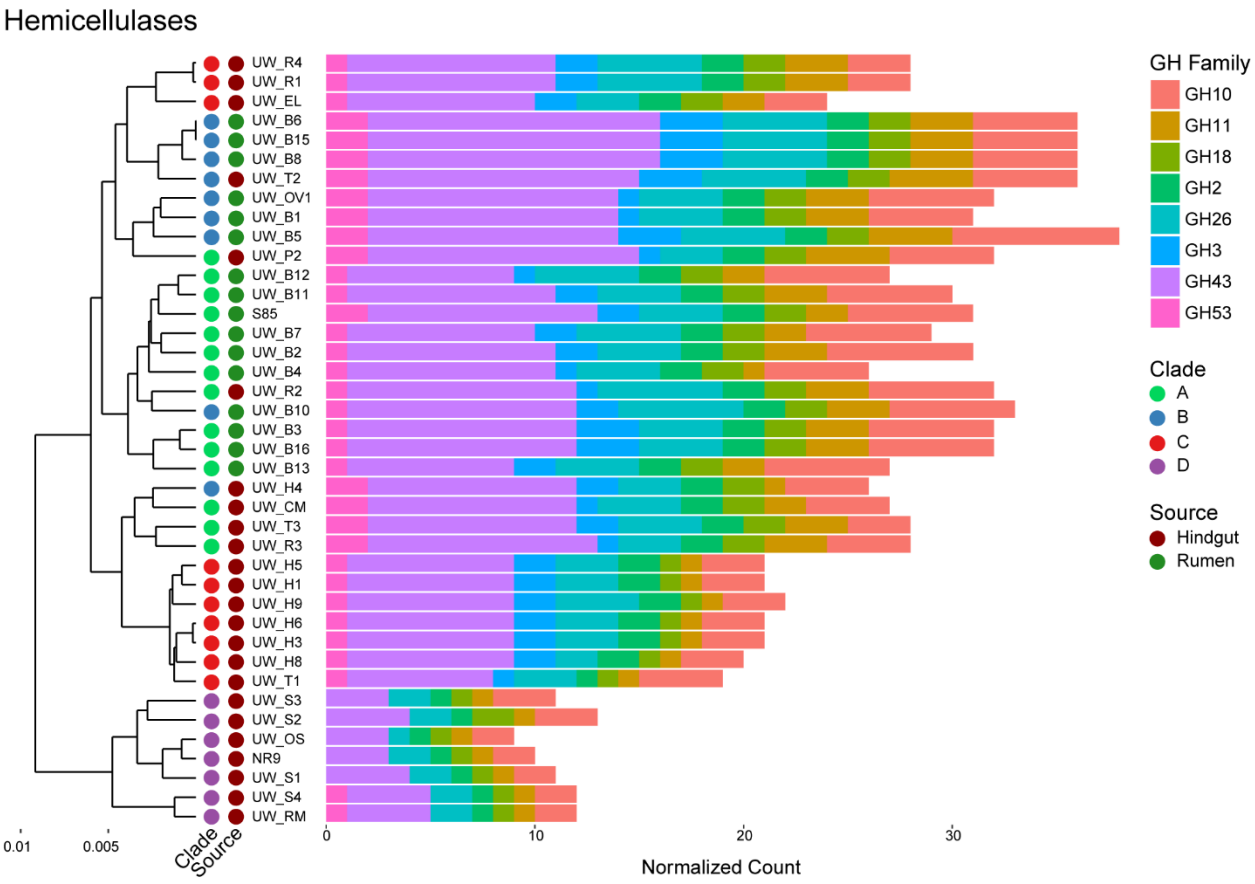


Figure 3.4: Distribution of hemicellulases among *Fibrobacter* strains. *Fibrobacter* strains are arranged according to the results of hierarchical clustering by relative abundance of all CAZyme family genes in their respective genomes. *Fibrobacter* strains are annotated with circles according to phylogenetic clade membership: clade A (green), clade B (blue), clade C (red), and clade D (violet), and isolation source: rumen (dark green) or hindgut (dark red). Horizontal bar plots represent the number of normalized gene counts in each glycoside hydrolase (GH) family.

families significantly more abundant in clade D genomes versus either clade A or B genomes, though these two families did not differ between clade D and clade C genomes.

Enrichment in CAZyme genes in *Fibrobacter* clades containing rumen phylotypes

Significantly greater numbers of total CAZyme genes as well as genes predicted to encode proteins annotated to the CAZyme classes CE and GH were observed at higher abundances in *F. succinogenes* clades A and B, which contain phylotypes that typically dominate

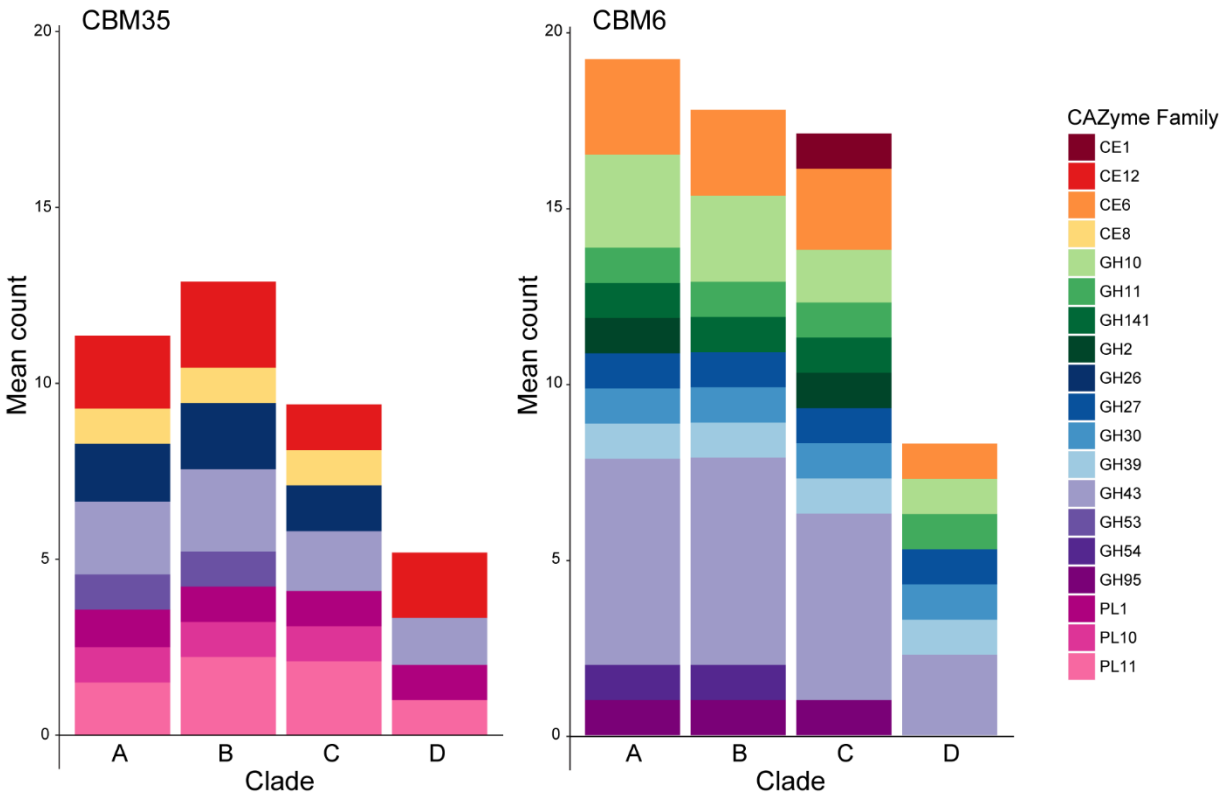


Figure 3.5: Distribution of catalytic CAZymes associated with CBM35 and CBM6 domains. The bar plots show the additive, mean gene counts of each catalytic CAZyme family present in the same coding sequence with either a CBM35 or CBM6 domain for each of the four *Fibrobacter* clades (A – D). The color coding of catalytic CAZyme families is consistent for both carbohydrate-binding module (CBM) domains.

in the rumen (17), compared to *F. succinogenes* clade C, which contains phylotypes that are uncommon in the rumen (Figure 3.7). The effect sizes of these differences were large, ranging from 1.2 to 4.8 standard deviations greater normalized gene counts in clade A or B genomes versus clade C genomes. *Fibrobacter* strains isolated from rumen samples generally had higher overall numbers of predicted cellulase and hemicellulase genes in their genomes. Upon fitting vectors representing the relative abundances of individual CAZyme families across the genomes to the dimensions of the NMDS analysis, several were identified that exhibited a strong positive association with areas of the scatterplot occupied by *Fibrobacter* strains of clade A and B (Figure 3.6). Significantly higher normalized genes counts for CAZyme families CE6, GH8,

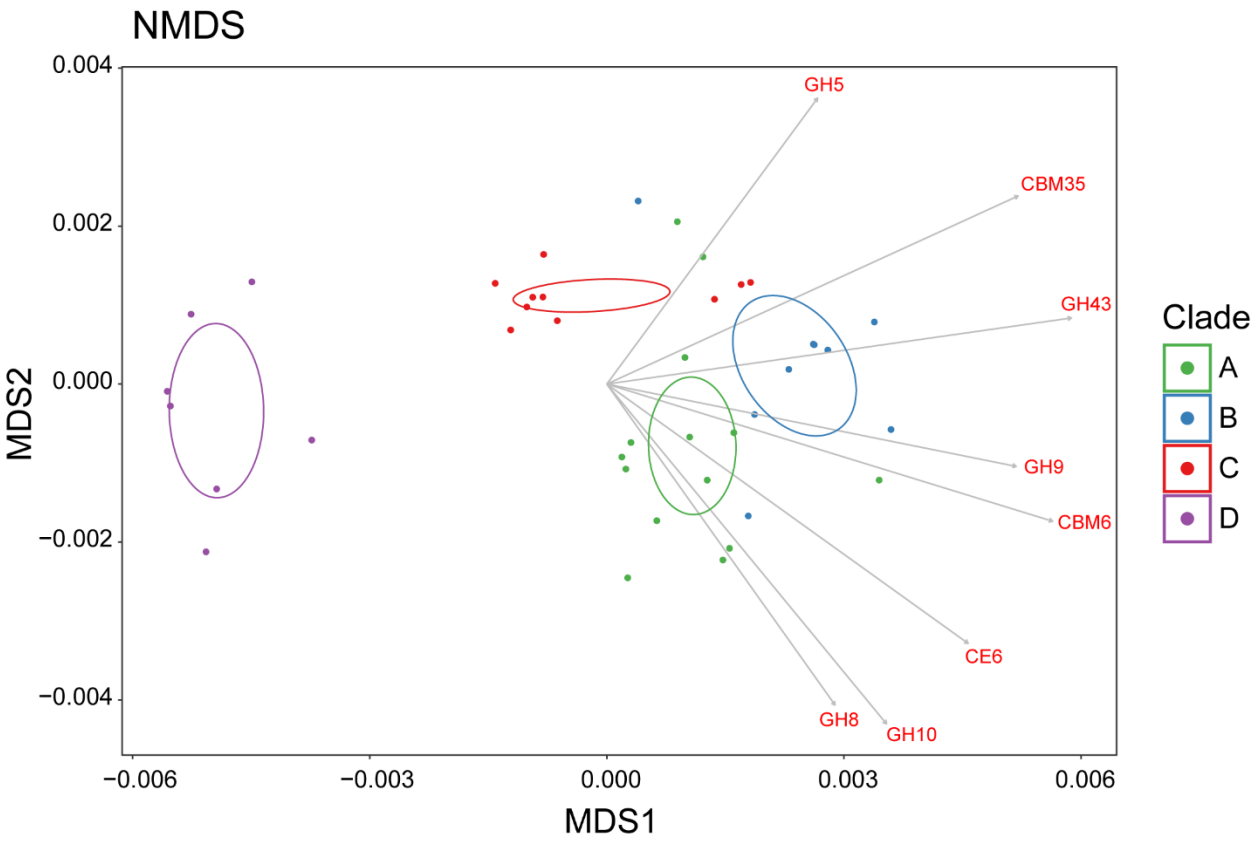


Figure 3.6: Scatterplot of 2-dimensional NMDS ordination based on the relative abundance of all CAZyme family genes in the *Fibrobacter* genomes. Individual points represent a single *Fibrobacter* strain, colored according to phylogenetic clade, clade A (green), clade B (blue), clade C (red), and clade D (violet). Open standard error ellipses (95% confidence interval) are plotted for each *Fibrobacter* clade, and colored accordingly. Vectors for CAZyme families, calculated by fitting the relative abundances of those CAZyme families in the *Fibrobacter* genomes to the dimensions of the NMDS, with significant fit to the ordination and a variance in median normalized count across the four clades are plotted with grey arrows and labeled accordingly in red

GH9, GH10, GH43, and GH45 were observed in clade A and clade B genomes compared with clade C (Figure 3.8). CAZyme family CBM35, which is often encoded by genes that also contain a catalytic CAZyme domain with demonstrated hemicellulase activity (18), was also significantly enriched in strains from clades A and B relative to those in clade C (Figure 3.8). Additional statistically significant differences in total CAZyme genes, total CEs and PLs were

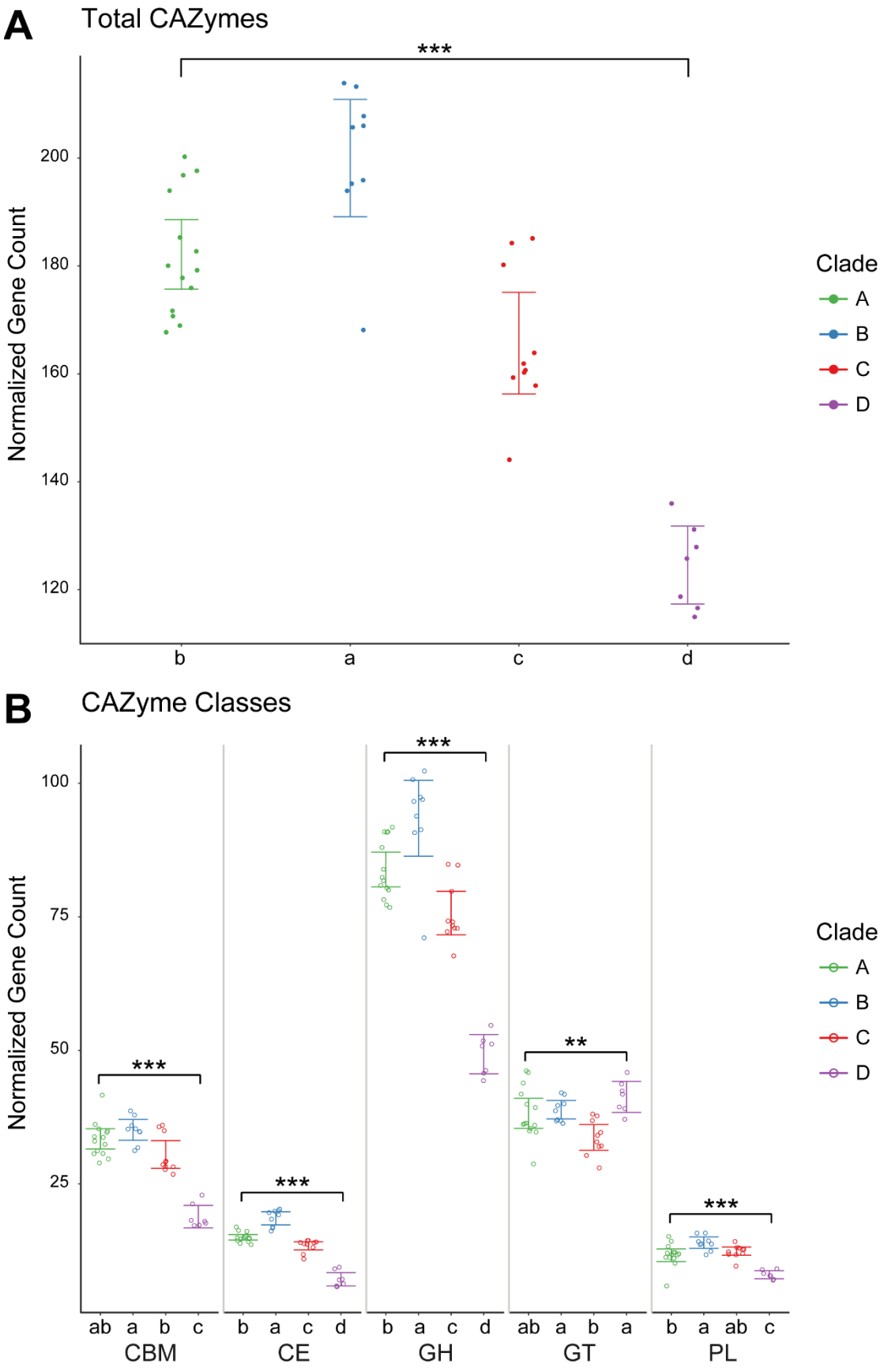

A
Total CAZymes
***
Normalized Gene Count
200
180
160
140
120
b
a
c
d
Clade
A
B
C
D
B
CAZyme Classes
Normalized Gene Count
100
75
50
25
***
***
***
**
***
ab a b c
b a c d
b a c d
ab a b a
b a ab c
CBM
CE
GH
GT
PL
Clade
A
B
C
D

Figure 3.7: Total CAZymes and CAZyme classes in *Fibrobacter* genomes. (A) Strip plot of normalized gene counts for total CAZymes in *Fibrobacter* genomes by phylogenetic clade. Each dot represents a single *Fibrobacter* strain, colored according to phylogenetic clade, clade A (green), clade B (blue), clade C (red), and clade D (violet). Error bars, colored according to clade, represent the 95% confidence interval for the mean normalized gene count for the clade. The bracket at the top of the plot represents the results of statistical testing for an overall difference among the clades with significance indicated as follows: $P < 0.05$ (·), $P < 0.01$ (*), $P < 0.001$ (**), $P < 0.0001$ (***) (ANOVA). Lowercase letters along the bottom of each plot represent significantly different groups based on pairwise statistical tests between individual clades (t-test, $P < 0.05$, Bonferroni's correction). (B) Strip plots of normalized gene counts for CAZyme classes. Each dot represents a single *Fibrobacter* strain, colored according to phylogenetic clade, clade A (green), clade B (blue), clade C (red), and clade D (violet). Error bars, colored according to clade, represent the 95% confidence interval for the mean normalized gene count for the clade. The plots are arranged horizontally by CAZyme class and labeled accordingly. Brackets represent the results of statistical testing for an overall difference among the clades for each CAZyme class with significance indicated as follows: $P < 0.05$ (·), $P < 0.01$ (*), $P < 0.001$ (**), $P < 0.0001$ (***) (ANOVA, Bonferroni's correction). Lowercase letters along the bottom of each plot represent significantly different groups based on pairwise statistical tests between individual clades (t-test, $P < 0.05$, Bonferroni's correction).

observed between clades A and B, with genomes from clade B having higher normalized gene counts in these categories (Figure 3.7).

Differences in carbon and nitrogen metabolism between rumen and hindgut phylotypes

All 40 of the *Fibrobacter* genomes analyzed contained genes for a complete glycolytic pathway through the formation of phosphoenolpyruvate (PEP), however the gene for pyruvate kinase was identified in the genomes of only 11 strains. These 11 strains were all originally isolated from hindgut fermenting hosts, had phylogenetic placements in either *Fibrobacter* clades C or D, and included *Fibrobacter* strain UWEL as well as all strains previously classified (17) to phylotypes Fs V and Fi I. All of the genomes contained a highly conserved gene predicted to encode PEP carboxykinase, which would allow for the formation of oxaloacetate and nucleoside triphosphate from PEP. Genes for pyruvate carboxylase, acetyl-coA synthetase, phosphate acetyltransferase, and acetate kinase were also identified in all strains, though the

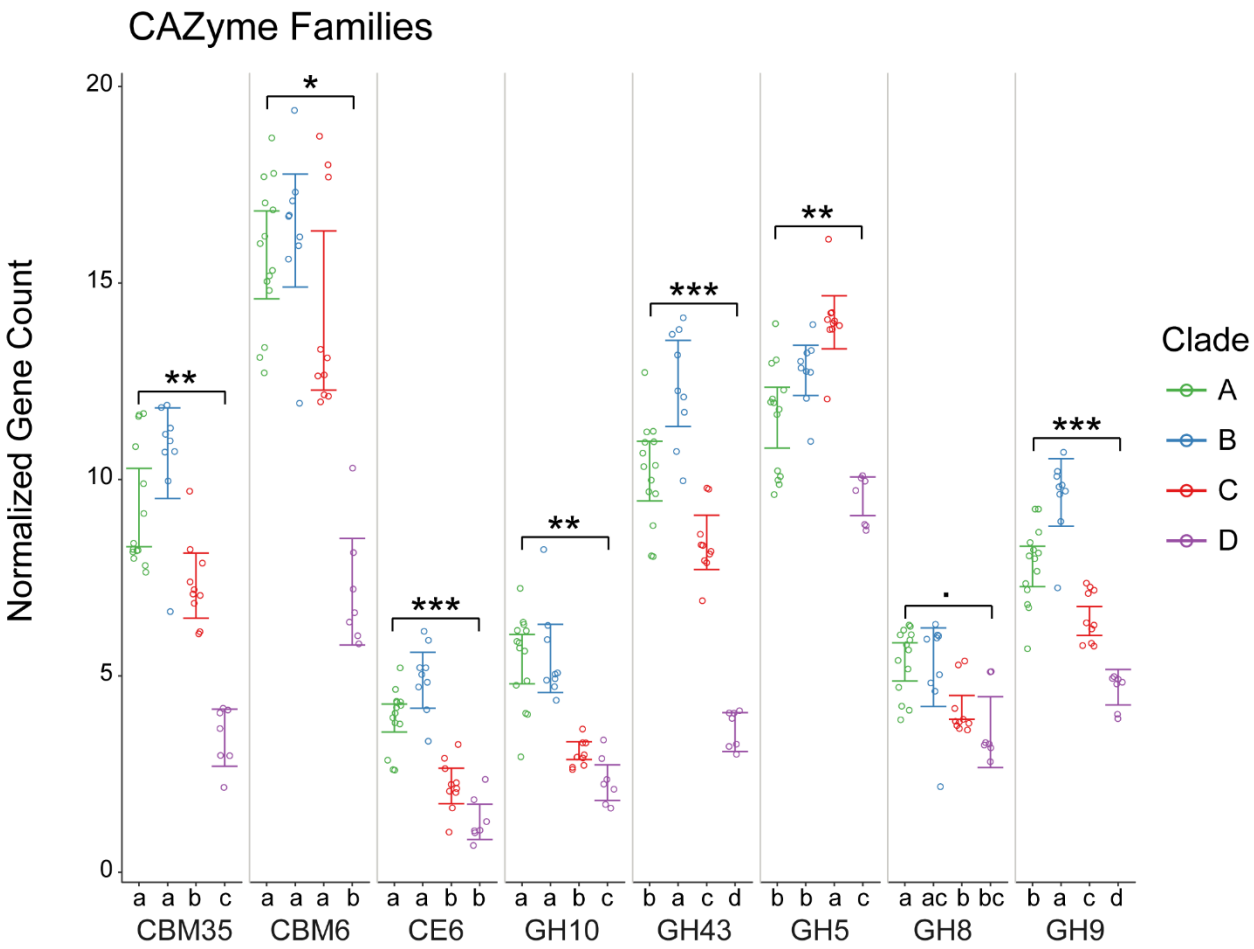
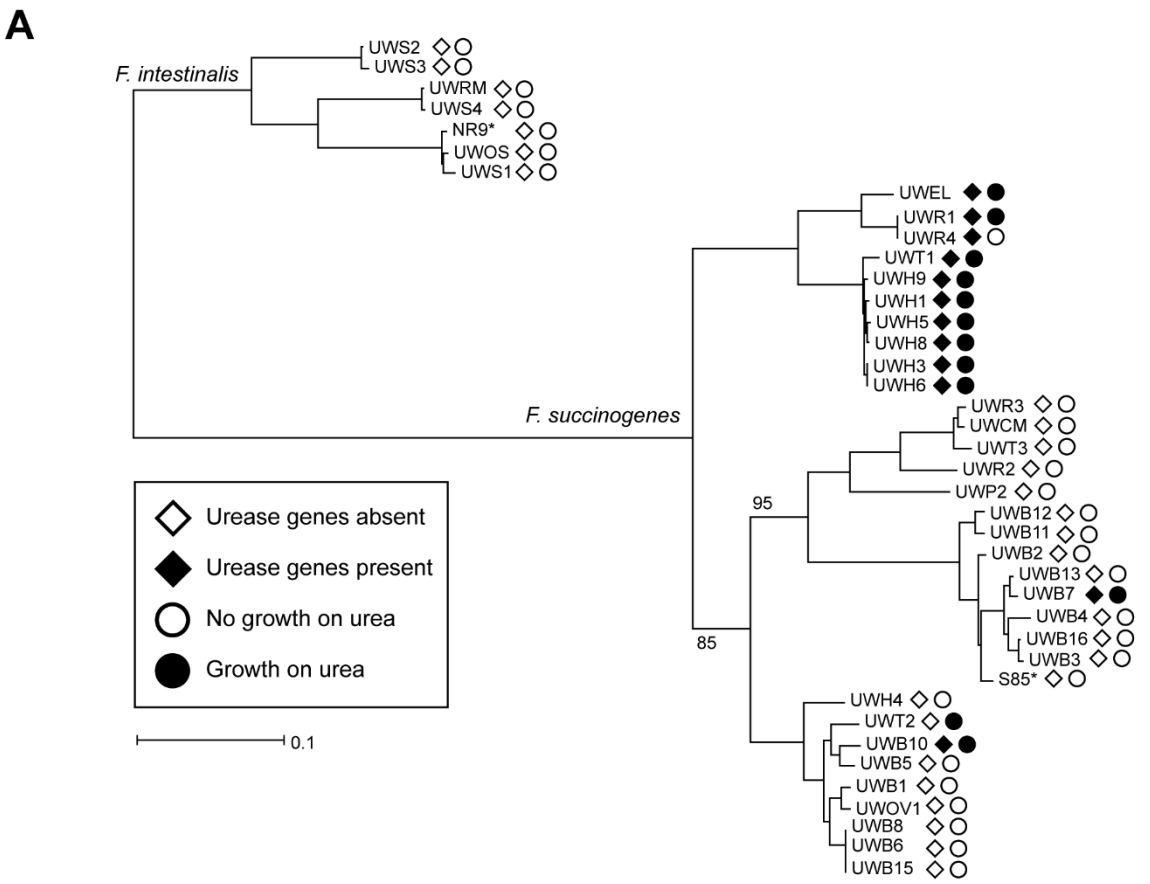


Figure 3.8: Strip plots of normalized gene counts for highly variable CAZyme families. Each dot represents a single *Fibrobacter* strain, colored according to phylogenetic clade, clade A (green), clade B (blue), clade C (red), and clade D (violet). Error bars, colored according to clade, represent the 95% confidence interval for the mean normalized gene count for the clade. The plots are arranged horizontally by CAZyme family and labeled accordingly. Brackets represent the results of statistical testing for an overall difference among the clades for each CAZyme family with significance indicated as follows: $P < 0.05$ (·), $P < 0.01$ (*), $P < 0.001$ (**), $P < 0.0001$ (***) (Kruskal-Wallis rank sum test, Holm's correction (46)). Lowercase letters along the bottom of each plot represent significantly different groups based on pairwise statistical tests between individual clades (Wilcoxon rank sum test, $P < 0.05$, Bonferroni's correction).

genes for formate C-acetyltransferase were missing from all 10 strains placed in clade C. The genomes suggest that, similar to other obligate anaerobes, all of the *Fibrobacter* strains use an incomplete, reductive TCA cycle, since genes for α-ketoglutarate dehydrogenase and succinyl-CoA synthetase are absent. All 40 *Fibrobacter* strains have a gene predicted to encode succinyl-

CoA:acetate CoA-transferase, and are therefore apparently capable of generating the essential metabolite succinyl-CoA via this enzyme. No strains had genes for enzymes of the oxidative branch of the pentose phosphate pathway, but all had genes predicted to encode enzymes of the non-oxidative branch. All 40 *Fibrobacter* strains were capable of growth on D-glucose and D-cellobiose, in addition to crystalline cellulose. None of the strains grew on L-arabinose, D-mannose, D-xylose, D-galactose, D-maltose, or L-rhamnose. Six strains, including the *F. succinogenes* type strain S85, exhibited growth on D-lactose, although no genetic basis for this phenotype could be readily identified (Table 3.3).

All strains were capable of growth on microcrystalline cellulose in media with essential growth factors and NH_4Cl, as the nitrogen source, lacking yeast extract or tryptone. All 40 *Fibrobacter* genomes had genes predicted to encode glutamate dehydrogenase and glutamine synthetase. During preliminary analysis an approximately 5.7 kb region predicted to encode two urease catalytic subunits and four urease accessory proteins was identified in several *Fibrobacter* clade C strains. This region was subsequently identified in 12 of the genomes, including all of the clade C genomes and two phylogenetically distinct rumen strains, UWB7 and UWB10, from clades A and B, respectively (Figure 3.9). Screening for growth on urea as the nitrogen source in media containing growth factors, but lacking yeast extract or tryptone confirmed that 11 of the 12 *Fibrobacter* strains containing the urease genes were capable of using urea as the primary source of nitrogen (Table 3.3). Moreover, all strains except one, UWT2, failed to grow in the media with urea as the sole nitrogen source further supporting the presence of a functional urease in strains containing the predicted urease genes.


A
F. intestinalis
UWS2
UWS3
UWRM
UWS4
NR9*
UWOS
UWS1
F. succinogenes
UWEL
UWR1
UWR4
UWT1
UWH9
UWH1
UWH5
UWH8
UWH3
UWH6
UWR3
UWCM
UWT3
UWR2
UWP2
95
UWB12
UWB11
UWB2
UWB13
UWB7
UWB4
UWB16
UWB3
S85*
85
UWH4
UWT2
UWB10
UWB5
UWB1
UWOV1
UWB8
UWB6
UWB15
Urease genes absent
Urease genes present
No growth on urea
Growth on urea
0.1

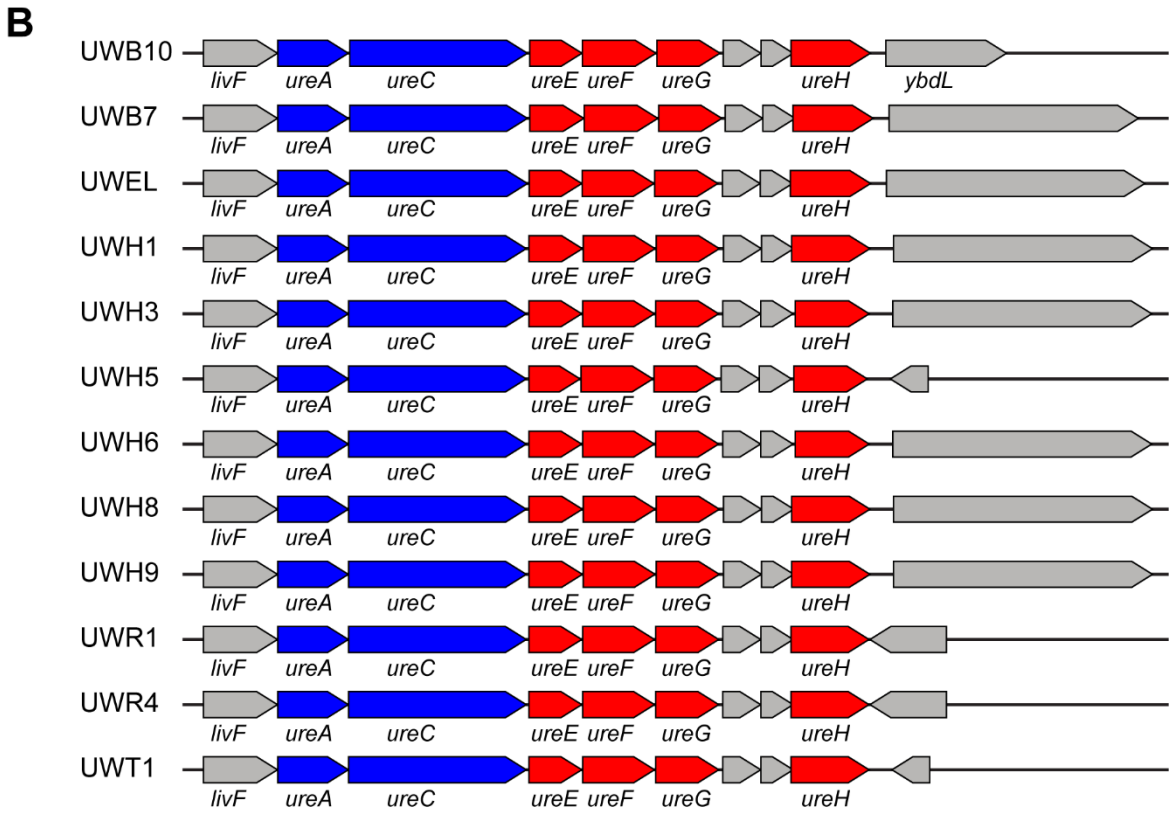

B
UWB10
UWB7
UWEL
UWH1
UWH3
UWH5
UWH6
UWH8
UWH9
UWR1
UWR4
UWT1
livF
ureA
ureC
ureE
ureF
ureG
ureH
ybdL

Figure 3.9: Presence of urease in *Fibrobacter* strains. (a) Maximum likelihood phylogeny constructed from the concatenated alignments of 99 essential protein genes with strains labeled according to the presence of urease genes and growth on urea as a source of nitrogen. Open diamonds (◊) represent the absence of the urease genes and filled diamonds (♦) represent the presence of urease genes in the genome. Open circles (○) indicate that the strain failed to grow in media with urea as the nitrogen source and filled circles (●) indicate that the strain was able to grow in media with urea as the source of nitrogen. (b) Aligned architecture of the approximately 5 kb region predicted to encode urease catalytic subunits (blue) and urease accessory proteins (red) in 12 positive strains. Genes not annotated as urease components are colored grey. Genes without labels are annotated as hypothetical proteins.

3.4 Discussion

Substantial phylogenetic diversity among bacteria in the genus *Fibrobacter* has been reported, but functional explanations for much of this diversity have proven elusive (26, 28, 29). We attempted to address this knowledge gap using whole genome comparisons among the largest collection of *Fibrobacter* genomes from axenic cultures generated to date. Considerably more phylogenetic diversity was observed among the 40 *Fibrobacter* strains than we previously estimated using full-length 16S rRNA gene sequences (17). A total of 21 different species-level groups were identified among the 40 strains from whole genome average nucleotide identity (ANI) using a cutoff of 95%. We had previously reported 9 discrete phylotypes among these same strains using a cutoff of 97% nucleotide identity across the 16S rRNA gene (17). Our results are consistent with what has generally been reported regarding differences in diversity estimates from whole genome sequences compared to the 16S rRNA gene (36). It remains to be determined whether the 21 different species-level groups, based on 95% ANI, are truly representative of ecologically discrete populations, but it is clear from these results that whole-genome sequence-based surveys tracking *Fibrobacter* populations through space and time are needed in order to fully understand the ecology and diversity of these fiber-degrading bacteria in their native environments.

Evolutionary relationships among the *Fibrobacter* strains predicted from the concatenated alignment of 99 essential proteins are largely in agreement with what we and others have reported previously for the *Fibrobacter* genus (17, 29, 37). All available *Fibrobacter* strains are clearly resolved into two discrete lineages corresponding to the species *F. succinogenes* and *F. intestinalis*. Moreover, four major clades were readily identified in the phylogeny, including three (A - C) made up of *F. succinogenes* strains and one (D) containing all strains of *F. intestinalis*. All of the phylotype designations previously identified using only full-length 16S rRNA gene sequences were preserved in the essential gene phylogeny produced here. However, the higher resolution obtained from the essential gene phylogeny allowed for the detection of levels of phylogenetic diversity suggestive of ecologically discrete populations present within several of the previously reported *Fibrobacter* phylotypes (32). Placement of the *Fibrobacter* phylotypes with respect to each other were similar to what was predicted previously (17), except that strains from phylotype Fs IV are predicted to share a more recent common ancestor with strains of Fs I than was originally predicted from the 16S rRNA gene sequences alone.

In addition to being among of the most actively cellulolytic mesophilic bacteria known, *Fibrobacter* has attracted interest due to its unusual strategy for PCWP degradation (17, 38, 39). Unlike many other anaerobic cellulose-degrading bacteria, which are found in the class Clostridia, stains of *Fibrobacter* do not produce a cellulosome to facilitate PCWP degradation (18). Homologs of genes for scaffoldin and cohesin domains, essential cellulosomal components, were similarly absent from the *Fibrobacter* genomes analyzed here. Comparisons in CAZyme gene content do suggest that the general mechanism of PCWP degradation by these bacteria is conserved. Genes for the major families of cellulases (GH5, GH8, GH9, and GH45)

and hemicellulases (GH10, GH26, GH43) typical of previously described *Fibrobacter* genomes were present in all strains (18, 31). While cellulases tended to occur without any identifiable associated CBM domain, hemicellulase domains tended to co-occur with predicted CBM6 and CBM35 domains. An exception was the identification of a single gene in all 40 *Fibrobacter* genomes predicted to encode a CBM11 domain along with a GH51 domain. This fits the profile of the Cel51A protein, also known as EG2 and CelF, a major cellulose-binding endoglucanase found in the outer membrane of *F. succinogenes* S85 (40, 41). It is unclear why only one copy of this gene exists in these genomes, when the other major cellulase families are usually present in multiple copies. Cel51A accounts for a large fraction of the total endoglucanase activity produced by S85 (42, 43). Analysis of global gene expression in S85 identified the gene for this enzyme as being among the most highly expressed of all genes in this strain, and also indicated that expression was constitutive (Chapter 4). These results support a particularly important role for this protein in PCWP degradation by *Fibrobacter*. A unique feature of the *Fibrobacter* cellulase system is the absence of processive exocellulases from CAZyme families GH6, GH7, and GH48 (31, 39). These CAZyme families were similarly not identified in the *Fibrobacter* genomes analyzed here. The distinctive mechanism of PCWP degradation by *Fibrobacter* likely also involves fibro-slime domain-containing proteins. These proteins have not been identified in any other bacteria and are distantly related to proteins produced by the slime mold *Dictyostelium discoideum* (35). Fibro-slime domain-containing proteins have been shown to bind cellulose and were originally identified in cellulose adherence-defective mutants of *F. succinogenes* S85 (44). Genes predicted to encode fibro-slime domains were identified in all 40 *Fibrobacter* genomes analyzed here, but the number of genes varied considerably ranging from 3 - 10 total genes. It's unclear exactly what role these proteins play in PCWP degradation other than assisting in

adherence, but an interesting hypothesis is that they may be involved in a form of gliding motility in *Fibrobacter* (18).

Despite sharing a similar general mechanism for PCWP degradation, substantial differences were observed in the numbers of total genes annotated to CAZyme families among phylogenetic groups. The most extreme example of this involved comparisons among representatives of the two *Fibrobacter* species. Strains from all three phylogenetic clades of *F. succinogenes* exhibited considerable enrichments in total CAZyme genes as well as CAZyme classes involved in PCWP degradation compared to strains of *F. intestinalis*. This enrichment in *F. succinogenes* genomes extended to several individual CAZyme families, particularly GHs. We have previously shown that phylotypes of *F. succinogenes* tend to predominate in the gastrointestinal tracts of ruminants and large hindgut fermenting mammals (17). A higher rate of fiber degradation by the gut microbiota in these hosts may provide a selective pressure to increase the capacity for enzyme production by these bacteria, explaining the general increases across CAZyme family genes observed in *F. succinogenes* relative to *F. intestinalis*. This selection would likely not be limited to *Fibrobacter*, and it has been reported that rumen strains of other polysaccharide-degrading organisms generally have higher proportions of CAZymes than their close relatives from less fibrolytic microbial communities (45–47). Although beneficial in highly fibrolytic environments like the rumen, the increased cost associated with synthesizing these extra enzymes might be detrimental in environments where the rate of fiber degradation is slower. This could explain why *F. intestinalis*, with its more streamlined CAZyme profile, is more abundant in the guts of animals with less fibrolytic microbiotas, *e.g.* pigs, apes, and rats. While screening for qualitative differences in the utilization of different carbohydrates among *Fibrobacter* strains has yielded little predictive information connecting

phylogeny to ecology of these organisms, the availability of axenic cultures of the strains analyzed here provide an opportunity to more rigorously test this hypothesis by screening for potential quantitative differences in their rates of fiber degradation.

Differences in CAZyme gene content were not limited to comparison between *Fibrobacter* species, but extended to comparisons among the three major phylogenetic clades of *F. succinogenes* (A - C). A similar general increase in the total number of CAZymes in clades A and B relative to clade C was observed, although not to the degree as was found for *F. succinogenes* versus *F. intestinalis*. Many of the major CAZyme families thought to be important in PCWP degradation by *Fibrobacter*, CE6, GH8, GH9, GH10, GH43, and GH45, were found at higher abundance in strains from clades A and B than they were in strains from clade C. *Fibrobacter* phylotypes in clades A and B predominate in the rumen (17). Fewer genes for enzymes in major *Fibrobacter* cellulase and hemicellulase families may contribute to the low relative abundance of phylotypes from *F. succinogenes* clade C in the rumen.

Some evidence for an effect of specific substrates on CAZyme gene content among the three major clades of *F. succinogenes* was also observed. Disparities in the total number of PLs in the genomes of strains from clades A, B, and C were less dramatic than they were for other CAZyme classes, and no difference was observed in total PL abundance between clades A and C. Moreover, a single CBM77 domain was commonly predicted to occur in strains of clade C, as well as several strains of *F. intestinalis*, but was not identified in any strains of clades A and B. PLs cleave uronic acid containing polysaccharides, which make up the backbone of pectin, and the only known activity of CBM77 is binding pectin (48). *Fibrobacter* strains solubilize pectin from intact forages, but there are conflicting reports regarding their ability to utilize this substrate for growth (49–51). The disparities in CAZymes predicted to target pectin among the

phylotypes compared here suggest that the ability to depolymerize this substrate may be of relatively greater importance for these bacteria in the hindgut than it is in the rumen. In the rumen, degradation of pectins occurs relatively rapidly, and involves diverse species of bacteria as well as anaerobic fungi (52–56). As a result, the pressure on rumen *Fibrobacter* phylotypes to degrade pectic polysaccharides themselves in order to access cellulose fibers may be less intense than it is for *Fibrobacter* phylotypes adapted to the hindgut.

Although phenotypic differences among the *Fibrobacter* isolates investigated here were few, genomic and functional evidence for some diversity of carbon and nitrogen metabolic capabilities between rumen and hindgut phylotypes was observed. Previous comparisons of Fibrobacteres genomes revealed variation in the presence of the gene for pyruvate kinase (31). We observed a similar result in our genome comparison. This gene was absent from all strains except for those closely related to the *F. intestinalis* type strain and most strains from clade C. An explanation for this variation is unclear, although strains with this gene tended to originate from hindgut fermenting herbivores. The gene encoding pyruvate formate lyase was also absent from strains in clade C. The lack of this gene could explain why formate production was not detected in the analysis of fermentation products produced by these strains (17). Variation in the ability to grow on D-lactose was also observed among the *Fibrobacter* strains analyzed here, although no clear genetic explanation for this phenotype was identified. Growth on lactose by *F. succinogenes* S85 has previously been attributed to extracellular β-galactosidase activity (57, 58). All 40 *Fibrobacter* genomes contained genes predicted to encode β-galactosidase, however we cannot rule out potential differences in expression that might explain the phenotypic diversity observed. It is worth noting that 5 of the 6 strains demonstrating growth on D-lactose were of

rumen origin, as it has recently been hypothesized that the ability of some *Fibrobacter* strains to utilize lactose provides them with a selective advantage in the developing calf rumen (58).

Genomic comparisons among the *Fibrobacter* strains revealed the presence of a contiguous stretch of genes predicted to encode for two catalytic subunits and accessory proteins for urease. As a result of this observation, the *Fibrobacter* strains were screened for growth on urea as the sole nitrogen source. All but one of the strains possessing the urease genes were capable of growth using urea as a source of nitrogen. *Fibrobacter* strains capable of utilizing urea were primarily from hindgut fermenting herbivores and found in clade C of the essential gene phylogeny, although 2 of the 16 rumen strains showed this ability. Urea makes up a larger fraction of the nitrogen available to microorganisms in the hindgut than it does in the rumen because a portion of the urea synthesized by the host is transferred from the blood into the intestinal lumen of the lower gut (59, 60). This explains why this trait was common in strains of *F. succinogenes* isolated from large hindgut fermenting herbivores, but much less so in strains of *F. succinogenes* isolated from the rumen. It remains to be determined how *Fibrobacter* strain UWT2 was able to grow readily in media with urea as the sole source of nitrogen when none of the urease genes were identified in the genome of this strain, but further investigation is warranted to determine if this strain possesses a unique mechanism for accessing nitrogen from urea.

In this study, we extended our previous work describing the isolation and ecological distribution of diverse *Fibrobacter* strains by performing the most extensive comparative analysis of *Fibrobacter* genomes available to date. Clear differences in CAZyme gene content were observed among these strains, and these differences were strongly correlated with a whole genome phylogeny based on amino acid sequences of essential genes. These results suggest that

CAZymes are particularly active targets for genome evolution in *Fibrobacter*. Comparative genomics identified variation in carbon and nitrogen metabolism among phylotypes that likely contributes to the ecology of these bacteria as well. The 38 novel *Fibrobacter* genomes generated here as well as the availability of axenic cultures for these strains provide valuable tools for testing specific hypotheses in future work aimed at unraveling the mechanism of PCWP degradation by *Fibrobacter* as well as the contribution of these bacteria to a healthy and efficient herbivore gut microbiome.

3.5 Materials & Methods

Fibrobacter strains and culture conditions

Isolation of the *Fibrobacter* strains analyzed in this study, along with their general characteristics such as major fermentation acids produced, has been described previously (17). Metadata for the samples from which they were isolated can be found in Table 3.1. *Fibrobacter* were typically cultivated at 39° C in an anaerobic atmosphere consisting of 5% H_2 + 20% CO_2, with the balance N_2. The growth medium used was a slightly modified version of the formula originally described by Scott and Dehority with 0.5% (wt vol^{-1}) Sigmacell 20 (Sigma-Aldrich, St. Louis, MO) as the primary carbon source (61). A detailed description of the formula can be found in Table 2.3. The same media formulation was used when testing for growth on different soluble sugars, except that cellulose was replaced with the soluble sugar being tested. Soluble sugars were included in the medium at a final concentration of 20 mM, with the exception of maltose which was included at a final concentration of 10 mM. Screening for growth with 5 mM urea as the sole nitrogen source was performed using the aforementioned media with microcrystalline cellulose as the carbon source and the following modifications: decreased L-

Cysteine-HCl from 1 g L^{-1} to 0.1 g L^{-1}, no NH_4Cl, no tryptone, and no yeast extract. All strains were capable of growth in this version of the medium with 10 mM NH_4Cl as the sole nitrogen source. Positive growth was defined as no reduction in turbidity upon outgrowth after at least two sequential 1% (vol vol^{-1}) transfers.

Genomic DNA extraction

High quality genomic DNA was obtained using the DOE Joint Genome Institute (JGI) method for Bacterial genomic DNA isolation using CTAB (http://jgi.doe.gov/user-program-info/pmo-overview/protocols-sample-preparation-information/). Briefly, cells from 10 mL of broth culture were recovered via centrifugation and resuspended in 740 µL TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH = 8.0). 40 µL 10% SDS and 8 µL proteinase K (10 mg mL^{-1}) were added followed by incubation at 56° C for 1 h. Next, 100 µL 5 M NaCl and 100µL CTAB/NaCl (275 mM hexadecyltrimethyl ammonium bromide, 700 mM NaCl) solution was added followed by incubation at 65° C for 10 min. The DNA was then sequentially extracted with chloroform, phenol:chloroform:isoamyl alcohol, chloroform, and precipitated with isopropanol. The DNA was resuspended in TE buffer and contaminating RNA was removed with RNase I (Epicentre, Madison, WI). gDNA was quantified using the BR dsDNA assay kit and Qubit® Fluorometer (Invitrogen, Carlsbad, CA) and checked for RNA contamination by agarose gel electrophoresis.

Genome sequencing, assembly and annotation

Whole-genome sequences were generated from either 300 or 550 bp insert standard Illumina shotgun libraries (Illumina, San Diego, CA) sequenced using either the Illumina HiSeq or MiSeq platforms at the UW Madison Biotechnology Center and DOE JGI (Walnut Creek,

CA). Whole-genome sequences for five isolates were generated at the DOE JGI from SMRTbell™ libraries using the PacBio RS platform (Pacific BioSciences, Menlo Park, CA). Raw Illumina reads were quality filtered per JGI SOP 1061 using BBTools (http://bbtools.jgi.doe.gov) and assembled using SPAdes, v3.9.0; —phred–offset 33 —cov–cutoff auto –m 40 —careful –k 25,55,95 —12 (62). Contigs < 1000 bp in length were discarded. Raw PacBio reads were assembled using HGAP, v2.3.0_p5; protocol version=2.3.0 method=RS HGAP Assembly.3,smrtpipe.py v1.87.139483 (63). Genome assemblies were assessed for quality and completeness using CheckM v1.0.11 (64). General genomic features were identified and annotated using Prokka, v1.12 (65). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations were determined using the BlastKOALA online annotation tool (http://www.kegg.jp/blastkoala/) for K number assignment (66). Carbohydrate-active enzyme (CAZyme) annotation was performed using HMMER, v3.1 (67) and the dbCAN database, v6 (34). CAZyme annotations with e-value > 1×10^{-18} and coverage < 0.35 were discarded, as recommended for bacteria (http://csbl.bmb.uga.edu/dbCAN/).

Comparative and phylogenetic analyses

Genome-wide average nucleotide identity (ANI) of shared genes was calculated using the *ani.rb* script from the Enveomics Collection (33). A distance matrix containing the ANI distances for all pairwise comparisons between *Fibrobacter* strains was used as to perform complete-linkage hierarchical clustering using the R function *hclust*. Cluster membership was determined using a cutoff of 0.05, corresponding to 95% ANI, typical of species-level differentiation for bacteria (32). A phylogenetic model of the evolutionary relationships among the strains was constructed from concatenated alignments of 99 essential proteins identified in all

40 *Fibrobacter* genomes using the *HMM.essential.rb* script from the Enveomics Collection (33). Amino acid sequences for each of the 99 essential proteins were aligned using Clustal Omega v1.2.4 (68) prior to concatenation by strain. The maximum likelihood phylogeny was inferred from the concatenated alignment using RAxML v8.2.11 (69, 70). The PROTGAMMAGTR substitution model was used with rapid bootstrapping. The resulting phylogenetic tree was midpoint rooted and visualized using Dendroscope v3.5.9 (71).

Statistical analysis of CAZyme gene content

CAZyme gene annotations for all 40 *Fibrobacter* strains were combined into a single data frame in R v3.4.3 (https://www.r-project.org). CAZyme gene content was normalized to control for differences in genome size among the strains by dividing the counts by the total number of genes identified in the corresponding genome to calculate the relative abundance for each CAZyme family. Normalized counts were calculated by multiplying the relative abundance by 2,553, the number of genes identified in the smallest *Fibrobacter* genome (UWS4), and rounding to the nearest integer. Normalized counts for total CAZyme genes and the CAZyme classes were assessed for normality and homoscedasticity by visually examining histograms and performing Levene's test of equality of variances using the function *levene.test* from the "lawstat" package. The 40 *Fibrobacter* strains were grouped into 4 major clades (A - D) based on the results of the maximum likelihood phylogeny inferred from essential protein sequences. Statistical differences in total CAZyme genes, and for each CAZyme class present in the *Fibrobacter* genomes, among the four major clades was assessed by performing ANOVA using the normalized counts. The effect size (eta squared) indicating the proportion of the total variance in the dependent variable, normalized count, explained by the independent variable,

Fibrobacter clade, was calculated using the R function *etaSquared* from the "lsr" package. Failure to reject the null hypothesis of no difference in normalized counts among the clades justified proceeding with pairwise t-tests which were performed using the R function *pairwise.t.test* with Bonferroni's correction to adjust for multiple comparisons and the pool.sd parameter set to false. Effect sizes for pairwise tests were assessed by calculating Glass's delta, the mean difference between group 1 and group 2 divided by the standard deviation of the whole sample, using the R function *cohen.d* (pooled = F) from the "effsize" package. Statistical differences in normalized counts for individual CAZyme families among the *Fibrobacter* clades were investigated using nonparametric tests by first testing for an overall difference using the Kruskal-Wallis rank sum test and following up with pairwise Wilcoxon rank sum tests when appropriate. P-values resulting from the 72 Kruskal-Wallis rank sum tests, one for each CAZyme family identified among the 40 *Fibrobacter* strains, were adjusted for multiple comparisons using the R function *p.adjust*, which uses Holm's correction method by default (72). P-values resulting from the pairwise Wilcoxon rank sum tests were adjusted using Bonferroni's correction.

Hierarchical clustering of the 40 *Fibrobacter* strains according to CAZyme gene content was performed by calculating the Euclidean distances among the strains from the matrix of CAZyme family relative abundances in the respective genomes followed by complete-linkage clustering using the R functions *dist* and *hclust*, respectively. A dendrogram of the results was constructed using functions from the R packages "ggdendro" and "ggplot2". Nonmetric multidimensional scaling (NMDS) analysis of the strains based on CAZyme gene content was performed using the CAZyme family relative abundances as input for the *metaMDS* function from the R package "vegan" (73). Vectors corresponding to the relative abundances of

individual CAZyme families were fit to the two dimensions of the NMDS using the “vegan” function *envfit*. CAZyme families present in all 40 *Fibrobacter* genomes were identified and used to construct a clustered heatmap of individual deviations in CAZyme family relative abundance from the mean for each strain and conserved CAZyme family using the function *pheatmap* from the R package “pheatmap”. Plots were generated using the R package “ggplot2”. All means are reported with their respective standard deviations.

Sequence accession numbers

Genome assemblies have been submitted to the National Center for Biotechnology Information and deposited in the GenBank and RefSeq databases.

3.6 References

1. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. 2008. Evolution of mammals and their gut microbes. Science 320:1647–1651.
2. Mackie RI. 2002. Mutualistic fermentative digestion in the gastrointestinal tract: diversity and evolution. Integr Comp Biol 42:319–326.
3. Loqué D, Scheller HV, Pauly M. 2015. Engineering of plant cell walls for enhanced biofuel production. Current Opinion in Plant Biology 25:151–161.
4. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol 6:121–131.
5. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37:D233-238.
6. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490-495.
7. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels 6:41.
8. Christopherson MR, Dawson JA, Stevenson DM, Cunningham AC, Bramhacharya S, Weimer PJ, Kendziorski C, Suen G. 2014. Unique aspects of fiber degradation by the ruminal ethanologen *Ruminococcus albus* 7 revealed by physiological and transcriptomic analysis. BMC Genomics 15:1066.
9. Brumm P, Mead D, Boyum J, Drinkwater C, Deneke J, Gowda K, Stevenson D, Weimer P. 2011. Functional annotation of *Fibrobacter succinogenes* S85 carbohydrate active enzymes. Appl Biochem Biotechnol 163:649–657.
10. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Doering DT, Adams AS, Blodgett JAV, Clardy J, Raffa KF, Fox BG, Currie CR. 2014. Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose. Appl Environ Microbiol 80:4692–4701.
11. Hemme CL, Mouttaki H, Lee Y-J, Zhang G, Goodwin L, Lucas S, Copeland A, Lapidus A, Glavina del Rio T, Tice H, Saunders E, Brettin T, Detter JC, Han CS, Pitluck S, Land ML, Hauser LJ, Kyrpides N, Mikhailova N, He Z, Wu L, Van Nostrand JD, Henrissat B, He Q, Lawson PA, Tanner RS, Lynd LR, Wiegel J, Fields MW, Arkin AP, Schadt CW, Stevenson BS, McInerney MJ, Yang Y, Dong H, Xing D, Ren N, Wang A, Huhnke RL, Mielenz JR, Ding S-Y, Himmel ME, Taghavi S, van der Lelie D, Rubin EM, Zhou J. 2010. Sequencing of multiple clostridial genomes related to biomass conversion and biofuel production. J Bacteriol 192:6494–6496.
12. Weimer PJ. 1996. Why don't ruminal bacteria digest cellulose faster? J Dairy Sci 79:1496–1502.
13. Leschine SB. 1995. Cellulose degradation in anaerobic environments. Annu Rev Microbiol 49:399–426.
14. O'Sullivan AC. 1997. Cellulose: the structure slowly unravels. Cellulose 4:173–207.
15. Weimer PJ. 1992. Cellulose degradation by ruminal microorganisms. Critical Reviews in Biotechnology 12:189–223.

16. Henderson G, Cox F, Ganesh S, Jonker A, Young W, Collaborators GRC, Abecia L, Angarita E, Aravena P, Arenas GN, Ariza C, Attwood GT, Avila JM, Avila-Stagno J, Bannink A, Barahona R, Batistotti M, Bertelsen MF, Brown-Kav A, Carvajal AM, Cersosimo L, Chaves AV, Church J, Clipson N, Cobos-Peralta MA, Cookson AL, Cravero S, Carballo OC, Crosley K, Cruz G, Cucchi MC, Barra R de la, Menezes ABD, Detmann E, Dieho K, Dijkstra J, Reis WLS dos, Dugan MER, Ebrahimi SH, Eythórsdóttir E, Fon FN, Fraga M, Franco F, Friedeman C, Fukuma N, Gagić D, Gangnat I, Grilli DJ, Guan LL, Miri VH, Hernandez-Sanabria E, Gomez AXI, Isah OA, Ishaq S, Jami E, Jelincic J, Kantanen J, Kelly WJ, Kim S-H, Klieve A, Kobayashi Y, Koike S, Kopecny J, Kristensen TN, Krizsan SJ, LaChance H, Lachman M, Lamberson WR, Lambie S, Lassen J, Leahy SC, Lee S-S, Leiber F, Lewis E, Lin B, Lira R, Lund P, Macipe E, Mamuad LL, Mantovani HC, Marcoppido GA, Márquez C, Martin C, Martinez G, Martinez ME, Mayorga OL, McAllister TA, McSweeney C, Mestre L, Minnee E, Mitsumori M, Mizrahi I, Molina I, Muenger A, Muñoz C, Murovec B, Newbold J, Nsereko V, O'Donovan M, Okunade S, O'Neill B, Ospina S, Ouwerkerk D, Parra D, Pereira LGR, Pinares-Patiño C, Pope PB, Poulsen M, Rodehutscord M, Rodriguez T, Saito K, Sales F, Sauer C, Shingfield K, Shoji N, Simunek J, Stojanović-Radić Z, Stres B, Sun X, Swartz J, Tan ZL, Tapio I, Taxis TM, Tomkins N, Ungerfeld E, Valizadeh R, Adrichem P van, Hamme JV, Hoven WV, Waghorn G, Wallace RJ, Wang M, Waters SM, Keogh K, Witzig M, Wright A-DG, Yamano H, Yan T, Yáñez-Ruiz DR, Yeoman CJ, Zambrano R, Zeitz J, Zhou M, Zhou HW, Zou CX, Zunino P, 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.
17. Neumann AP, McCormick CA, Suen G. 2017. *Fibrobacter* communities in the gastrointestinal tracts of diverse hindgut-fermenting herbivores are distinct from those of the rumen. Environ Microbiol 19:3768–3783.
18. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D, Brumm PJ. 2011. The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. PLoS ONE 6:e18814.
19. Coen JA, Dehority BA. 1970. Degradation and utilization of hemicellulose from intact forages by pure cultures of rumen bacteria. Appl Microbiol 20:362–368.
20. Bryant MP, Robinson IM, Chu H. 1959. Observations on the nutrition of *Bacteroides succinogenes*—A ruminal cellulolytic bacterium. Journal of Dairy Science 42:1831–1847.
21. Scheifinger CC, Wolin MJ. 1973. Propionate formation from cellulose and soluble sugars by combined cultures of *Bacteroides succinogenes* and *Selenomonas ruminantium*. Appl Microbiol 26:789–795.
22. Blackburn TH, Hungate RE. 1963. Succinic acid turnover and propionate production in the bovine rumen. Appl Microbiol 11:132–135.
23. Bergman EN, Roe WE, Kon K. 1966. Quantitative aspects of propionate metabolism and gluconeogenesis in sheep. Am J Physiol 211:793–799.
24. Chaucheyras-Durand F, Masséglia S, Fonty G, Forano E. 2010. Influence of the composition of the cellulolytic flora on the development of hydrogenotrophic microorganisms, hydrogen utilization, and methane production in the rumens of gnotobiotically reared lambs. Appl Environ Microbiol 76:7931–7937.

25. Hungate RE. 1947. Studies on cellulose fermentation: III. The culture and isolation for cellulose-decomposing bacteria from the rumen of cattle. J Bacteriol 53:631–645.
26. Jewell KA, Scott JJ, Adams SM, Suen G. 2013. A phylogenetic analysis of the phylum Fibrobacteres. Syst Appl Microbiol 36:376–382.
27. Ransom-Jones E, McCarthy AJ, Haldenby S, Doonan J, McDonald JE. 2017. Lignocellulose-degrading microbial communities in landfill sites represent a repository of unexplored biomass-degrading diversity. mSphere 2.
28. Ransom-Jones E, Jones DL, Edwards A, McDonald JE. 2014. Distribution and diversity of members of the bacterial phylum Fibrobacteres in environments where cellulose degradation occurs. Syst Appl Microbiol 37:502–509.
29. Amann RI, Lin C, Key R, Montgomery L, Stahl DA. 1992. Diversity among *Fibrobacter* isolates: towards a phylogenetic classification. Systematic and Applied Microbiology 15:23–31.
30. Montgomery L, Flesher B, Stahl D. 1988. Transfer of *Bacteroides succinogenes* (Hungate) to *Fibrobacter* gen. nov. as *Fibrobacter succinogenes* comb. nov. and description of *Fibrobacter intestinalis* sp. nov. International Journal of Systematic and Evolutionary Microbiology 38:430–435.
31. Abdul Rahman N, Parks DH, Vanwonterghem I, Morrison M, Tyson GW, Hugenholtz P. 2015. A phylogenomic analysis of the bacterial phylum Fibrobacteres. Front Microbiol 6:1469.
32. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. 2015. Microbial species delineation using whole genome sequences. Nucleic Acids Res 43:6761–6771.
33. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. e1900v1. PeerJ Inc.
34. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445-451.
35. Jun H-S, Qi M, Gong J, Egbosimba EE, Forsberg CW. 2007. Outer membrane proteins of *Fibrobacter succinogenes* with potential roles in adhesion to cellulose and in cellulose digestion. J Bacteriol 189:6806–6815.
36. Rodriguez-R LM, Castro JC, Kyrpides NC, Cole JR, Tiedje JM, Konstantinidis KT. 2018. How much do rRNA gene surveys underestimate extant bacterial diversity? Appl Environ Microbiol 84.
37. Shinkai T, Ohji R, Matsumoto N, Kobayashi Y. 2009. Fibrolytic capabilities of ruminal bacterium *Fibrobacter succinogenes* in relation to its phylogenetic grouping. FEMS Microbiol Lett 294:183–190.
38. Burnet MC, Dohnalkova AC, Neumann AP, Lipton MS, Smith RD, Suen G, Callister SJ. 2015. Evaluating models of cellulose degradation by *Fibrobacter succinogenes* S85. PLOS ONE 10:e0143809.
39. Wilson DB. 2009. Evidence for a novel mechanism of microbial cellulose degradation. Cellulose 16:723–727.
40. Qi M, Jun H-S, Forsberg CW. 2007. Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. Appl Environ Microbiol 73:6098–6105.
41. Raut MP, Karunakaran E, Mukherjee J, Biggs CA, Wright PC. 2015. Influence of substrates on the surface characteristics and membrane proteome of *Fibrobacter succinogenes* S85. PLOS ONE 10:e0141197.

42. McGavin M, Forsberg CW. 1988. Isolation and characterization of endoglucanases 1 and 2 from *Bacteroides succinogenes* S85. J Bacteriol 170:2914–2922.
43. McGavin M, Forsberg CW. 1989. Catalytic and substrate-binding domains of endoglucanase 2 from *Bacteroides succinogenes*. J Bacteriol 171:3310–3315.
44. Gong J, Forsberg CW. 1989. Factors affecting adhesion of *Fibrobacter succinogenes* subsp. *succinogenes* S85 and adherence-defective mutants to cellulose. Appl Environ Microbiol 55:3039–3044.
45. Accetto T, Avguštin G. 2015. Polysaccharide utilization locus and CAZYme genome repertoires reveal diverse ecological adaptation of *Prevotella* species. Syst Appl Microbiol 38:453–461.
46. Mukhopadhya I, Moraïs S, Laverde-Gomez J, Sheridan PO, Walker AW, Kelly W, Klieve AV, Ouwerkerk D, Duncan SH, Louis P, Koropatkin N, Cockburn D, Kibler R, Cooper PJ, Sandoval C, Crost E, Juge N, Bayer EA, Flint HJ. 2018. Sporulation capability and amylosome conservation among diverse human colonic and rumen isolates of the keystone starch-degrader *Ruminococcus bromii*. Environ Microbiol 20:324–336.
47. Israeli-Ruimy V, Bule P, Jindou S, Dassa B, Moraïs S, Borovok I, Barak Y, Slutzki M, Hamberg Y, Cardoso V, Alves VD, Najmudin S, White BA, Flint HJ, Gilbert HJ, Lamed R, Fontes CMGA, Bayer EA. 2017. Complexity of the *Ruminococcus flavefaciens* FD-1 cellulosome reflects an expansion of family-related protein-protein interactions. Sci Rep 7:42355.
48. Venditto I, Luis AS, Rydahl M, Schückel J, Fernandes VO, Vidal-Melgosa S, Bule P, Goyal A, Pires VMR, Dourado CG, Ferreira LMA, Coutinho PM, Henrissat B, Knox JP, Baslé A, Najmudin S, Gilbert HJ, Willats WGT, Fontes CMGA. 2016. Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. PNAS 113:7136–7141.
49. Bryant MP, Doetsch RN. 1954. A study of actively cellulolytic rod-shaped bacteria of the bovine rumen. J Dairy Sci 37:1176–1183.
50. Osborne JM, Dehority BA. 1989. Synergism in degradation and utilization of intact forage cellulose, hemicellulose, and pectin by three pure cultures of ruminal bacteria. Appl Environ Microbiol 55:2247–2250.
51. Gradel CM, Dehority BA. 1972. Fermentation of isolated pectin and pectin from intact forages by pure cultures of rumen bacteria. Appl Microbiol 23:332–340.
52. Hatfield RD, Weimer PJ. 1995. Degradation characteristics of isolated and in situ cell wall lucerne pectic polysaccharides by mixed ruminal microbes. Journal of the Science of Food and Agriculture 69:185–196.
53. Paster BJ, Canale-Parola E. 1985. *Treponema saccharophilum* sp. nov., a large pectinolytic spirochete from the bovine rumen. Appl Environ Microbiol 50:212–219.
54. Dusková D, Marounek M. 2001. Fermentation of pectin and glucose, and activity of pectin-degrading enzymes in the rumen bacterium *Lachnospira multiparus*. Lett Appl Microbiol 33:159–163.
55. Marounek M, Dušková D. 2002. Metabolism of pectin in rumen bacteria *Butyrivibrio fibrisolvens* and *Prevotella ruminicola*. Letters in Applied Microbiology 29:429–433.
56. Kopecný J, Hodrová B. 1995. Pectinolytic enzymes of anaerobic fungi. Lett Appl Microbiol 20:312–316.

57. Javorsky P, Lee SF, Gibbins AM, Forsberg CW. 1990. Extracellular beta-galactosidase activity of a *Fibrobacter succinogenes* S85 mutant able to catabolize lactose. Appl Environ Microbiol 56:3657–3663.
58. Ghali I, Sofyan A, Ohmori H, Shinkai T, Mitsumori M. 2017. Diauxic growth of *Fibrobacter succinogenes* S85 on cellobiose and lactose. FEMS Microbiol Lett 364.
59. Kennedy PM, Milligan LP. 1980. The degradation and utilization of endogenous urea in the gastrointestinal tract of ruminants: a review. Can J Anim Sci 60:205–221.
60. Wegmann U, Louis P, Goesmann A, Henrissat B, Duncan SH, Flint HJ. 2014. Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota ('*Ruminococcus bicirculans*') reveals two chromosomes and a selective capacity to utilize plant glucans. Environ Microbiol 16:2879–2890.
61. Scott HW, Dehority BA. 1965. Vitamin requirements of several cellulolytic rumen bacteria. J Bacteriol 89:1169–1175.
62. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.
63. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569.
64. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.
65. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.
66. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731.
67. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195.
68. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.
69. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
70. Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21:456–463.
71. Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol 61:1061–1067.
72. Holm S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6:65–70.
73. Dixon P. 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science 14:927–930.

Chapter 4: Gene expression in *Fibrobacter succinogenes* S85 in response to growth rate and substrate

Authors and contributors:

Anthony Neumann: Designed the experiments, performed the experiments, analyzed the data, and wrote the manuscript

Dr. Paul Weimer: Designed the experiments, performed the experiments, and edited the manuscript

Dr. Garret Suen: Designed the experiments, and edited the manuscript

This work has been submitted for publication to *Biotechnology for Biofuels* (March 2018) as:

Neumann, A. P., Weimer, P. J., and Suen, G. A global analysis of gene expression in *Fibrobacter succinogenes* S85 grown on cellulose and soluble sugars at different growth rates.

Funding:

This work was supported by a traineeship from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number T32AI55397 to APN, a UW-Madison Richard M. Heins Wisconsin Distinguished Graduate Fellowship to APN, a U.S. Department of Energy Joint Genome Institute Community Science Program sequencing grant to APN and GS, and a US Department of Energy Biological and Environmental Research Early Career Research Program Award DE-SC0008104 to GS. This work was also supported by a USDA Agricultural Research Service (Washington, DC) CRIS project 5090-21000-024-00-D.

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The funders had no input in the design, execution, or interpretation of this work.

4.1 Abstract

Cellulose is the most abundant biological polymer on earth, making it an attractive substrate for the production of next-generation biofuels and commodity chemicals. However, the economics of cellulose utilization are currently unfavorable due to a lack of efficient methods for its hydrolysis. In this study, we examined the transcriptome of *F. succinogenes* S85 grown in continuous culture at several dilution rates on either cellulose, cellobiose, or glucose in order to gain a system-level understanding of cellulose degradation by this bacterium. Several patterns of gene expression were observed for the major cellulases produced by *F. succinogenes* S85. A large proportion of cellulase genes were constitutively expressed, including the gene encoding for Cel51A, the major cellulose-binding endoglucanase produced by this bacterium. Moreover, other cellulase genes displayed elevated expression during growth on cellulose relative to growth on soluble sugars. Growth rate had a strong effect on global gene expression, particularly with regard to genes predicted to encode carbohydrate-binding modules and glycoside hydrolases implicated in hemicellulose degradation. Expression of hemicellulase genes was tightly regulated, with these genes displaying elevated expression only during slow growth on soluble sugars. Clear differences in gene expression were also observed between adherent and planktonic populations within continuous cultures growing on cellulose. This work emphasizes the complexity of the fiber-degrading system utilized by *F. succinogenes* S85, and reinforces the complementary role of hemicellulases for accessing cellulose by these bacteria. We report for the first time evidence of global differences in gene expression between adherent and planktonic populations of an anaerobic bacterium growing on cellulose. Finally, our results also highlight the importance of controlling for growth rate in investigations of gene expression.

4.2 Introduction

Cellulose is the major structural polysaccharide in plant cell walls and is estimated to be the most abundant biological polymer on earth (1). As a result, cellulose is of fundamental importance to herbivorous animals, as well as an attractive substrate for the production of next-generation renewable fuels and commodity chemicals. However, the exploitation of cellulose as a source for energy is severely limited by the chemical nature of the cellulose molecule itself and the complex architecture of the plant cell wall. Cellulose is composed of repeating units of cellobiose, the β-(1-4)-linked disaccharide of glucose, a configuration that produces linear chains which can then assemble into higher order structures of increasing complexity such as: planar sheets, crystals, microfibrils, and fibers (2). Hydrogen bonding among adjacent chains and sheets reinforces the fiber, which is impermeable even to water. In the plant cell wall, these cellulose fibers are typically embedded in a matrix of hemicellulose, pectin, and lignin, further adding to their stability (3). Nevertheless, microorganisms have evolved mechanisms to access and depolymerize cellulose in order to utilize its constituent sugars for growth (4). Some of the most prolific of these cellulose-degrading microbes live in the herbivore gut (5).

Ruminants, animals with an expanded foregut for microbial fermentation of plant matter, are among the most successful herbivores known, and are of economic importance to humans throughout the world. The rumen microbiota has been extensively studied because of its superior cellulolytic ability and vital connection to the health and well-being of its host (6). Species of bacteria from the genera *Ruminococcus* and *Fibrobacter* are typically the most abundant cellulose-degrading microbes in the rumen, and are among the most actively cellulolytic organisms identified to date (7). *Fibrobacter succinogenes*, the major *Fibrobacter* species in the rumen, is especially adept at solubilizing highly ordered forms of cellulose (8, 9).

Moreover, strains of *F. succinogenes* have demonstrated a greater ability to digest cellulose from intact forages than other species of fibrolytic rumen bacteria (10). The *F. succinogenes* type strain, S85, was originally isolated from a bovine rumen fluid sample and has been utilized extensively as a model to study *Fibrobacter* physiology, as well as their unique mechanism of polysaccharide degradation (11–14). *F. succinogenes* S85 has attracted additional interest because it is one of the few known cellulolytic microorganisms to produce the valuable chemical succinic acid as its major fermentation product (15, 16).

Much progress has been made in our understanding of cellulolysis by *Fibrobacter*, but many important details of the process remain poorly understood. Like other anaerobic cellulose-degrading bacteria, *Fibrobacter* requires close physical contact with the substrate in order to achieve efficient hydrolysis and growth (17–19). However, unlike many clostridia, the most well-studied group of anaerobic cellulose-degrading bacteria, *Fibrobacter* does not produce cellulosomes to facilitate the process (20, 21). A Gram-negative cell envelope also differentiates *Fibrobacter* from the cellulolytic clostridia, and several unique outer membrane (OM) proteins have been implicated in cellulose-binding including fibro-slime domain containing proteins and a type IV pilin (22). The OM is apparently only loosely associated with the cell, and vesicles exhibiting hydrolytic activity have been observed originating from the OM (23, 24). The S85 genome has been predicted to encode 104 different glycoside hydrolases (GHs), representing 3.37% of all S85 genes, which is among the highest percentage of GHs in any bacterial species (25). Several of these enzymes have been purified and characterized, but they generally have limited or no activity against crystalline cellulose (26, 27).

Although many proteins have been implicated in the process of cellulose degradation by *Fibrobacter*, it is not well known how they interact or how their expression may be regulated.

Some reports have proposed constitutive expression of *Fibrobacter* cellulases while others have suggested tight regulation (28, 29). The goal of this work was to better understand the mechanism of cellulose degradation by *Fibrobacter* through genome-wide analysis of gene expression using RNA sequencing. In order to determine the effect of carbon source on gene expression, *F. succinogenes* S85 was cultured on cellulose, cellobiose, or glucose at several growth rates. Cells from cellulose-grown cultures were fractionated into cellulose-adherent and cellulose-planktonic populations and evaluated separately. The effect of growth rate was controlled by analyzing populations at steady-state during continuous culture. We hypothesized that genes important for digesting cellulose would be upregulated during growth on cellulose compared to growth on soluble sugars, and that patterns of co-expression would allow us to identify cellular processes that act synergistically in cellulose hydrolysis by *Fibrobacter*. Using this approach, we obtained evidence for tight control over the expression of hemicellulases by S85, in addition to identifying a subset of cellulases that were highly expressed under all conditions examined while a separate group of cellulases showed elevated expression during growth on cellulose. This work also confirmed the heterogeneous nature of bacterial populations cultured on an insoluble substrate, and reinforced the importance of controlling for the effect of growth rate when investigating microbial gene expression.

4.3 Results

General characteristics of *F. succinogenes* S85 continuous cultures

Summary statistics for the cultures investigated in this study, and the results of statistical analysis of fermentation products in steady-state culture supernatants are reported in Table 4.1. Dilution rates calculated at the time of sampling were all very close to their intended targets.

Table 4.1: Culture summary statistics

Substrate	Target D (h^{-1})	Actual D (h^{-1})	Succinate	Acetate	Formate	Total Soluble Sugar	SC20 consumed (g / L)
Sigmacell 20	0.018	0.018	15.56 ± 2.40	7.24 ± 1.00	0.52 ± 0.39	1.31 ± 0.08	1.83
Sigmacell 20	0.047	0.047	7.05 ± 0.97	3.11 ± 0.52	0.21 ± 0.38	0.56 ± 0.06	0.31
Sigmacell 20	0.068	0.068	9.56 ± 1.14	2.92 ± 0.47	0.09 ± 0.15	0.72 ± 0.05	0.74
Glucose	0.018	0.019	12.70 ± 0.41	5.01 ± 0.15	0.17 ± 0.03	0.81 ± 0.04	
Glucose	0.047	0.047	10.09 ± 0.49	4.39 ± 1.40	0.69 ± 0.46	1.45 ± 0.14	
Glucose	0.068	0.067	10.80 ± 0.35	3.86 ± 0.15	0.31 ± 0.08	0.98 ± 0.14	
Glucose	0.200	0.198	11.25 ± 1.68	3.79 ± 1.90	0.94 ± 0.55	0.84 ± 0.03	
Cellobiose	0.018	0.016	13.33 ± 0.99	4.66 ± 0.36	ND	0.97 ± 0.05	
Cellobiose	0.047	0.047	10.54 ± 0.44	3.36 ± 0.13	0.21 ± 0.11	1.19 ± 0.05	
Cellobiose	0.068	0.068	10.69 ± 0.36	3.58 ± 0.08	0.44 ± 0.09	1.01 ± 0.17	
Cellobiose	0.200	0.193	10.03 ± 0.49	3.32 ± 0.16	0.50 ± 0.08	0.82 ± 0.07	
		Growth rate (GR)	0.070	0.007			
ANOVA*		Substrate (S)	0.518	0.095			
		GR x S	0.002	1.18×10^{-6}			

Values of fermentation products and total soluble sugars are mean ± 1 standard deviation of three biological replicates
*Results reported for ANOVA of the two factors and interaction are p-values
D = Dilution rate
ND = not detected

Succinate, detected at concentrations ranging from 9.60 - 15.56 mM, was the major fermentation product regardless of substrate type or growth rate. Acetate was the only other fermentation product detected at concentrations > 1 mM. Traces (< 1 mM) of formate were detected in most cultures, but the concentrations detected, near the method's detection limit, were considered too low for reliable statistical analyses. Linear regression followed by ANOVA identified a significant interaction between carbon source and growth rate for both succinate ($P = 1.55 \times 10^{-3}$) and acetate ($P = 1.18 \times 10^{-6}$) production. Concentrations of both fermentation products fell much faster with increasing growth rate during growth on cellulose compared with growth on glucose or cellobiose; this was expected because cellulose utilization varied considerably with dilution rate, while cellobiose and glucose consumption was nearly complete at all dilution rates. In cellulose-fed continuous cultures, total cellulose consumption (1.828 g L^{-1}) was highest in the cultures growing at the lowest growth rate (D = 0.018 h^{-1}). In cellulose-grown cultures, the vast majority of the cells were adherent to cellulose (89.9%, 95.4% and 96.0% for D = 0.018, 0.047 and 0.068 h^{-1}, respectively, based on protein analysis of adherent and non-adherent cell pellets).

Regardless of substrate fed, all cultures contained 0.8 - 1.4 mM of total soluble sugars. These soluble sugar concentrations averaged 0.86 ± 0.40, 1.02 ± 0.29, and 1.00 ± 0.15 mM (mean ± s.d.) for cellulose, glucose, and cellobiose, respectively (means not different, $P > 0.05$). Separate analysis of reducing sugars indicated that the average degree of polymerization of the soluble sugars (calculated as mM total sugars / mM reducing sugars) averaged 1.41 across all substrates and did not differ across substrate.

Genome-wide expression by *F. succinogenes* S85 across conditions

Figure 4.1 shows the relationships among the transcriptomes of all samples according to the first two principal components of PCA. The first principal component (PC1) explained 44% of the total variance and generally differentiated the samples according to growth rate, with faster growing populations tending to have higher values of PC1 compared to those growing more slowly on the same substrate. PC1 also clearly differentiated the two subpopulations of cellulose cultures. Cellulose-adherent populations had higher values of PC1 compared to the corresponding cellulose-planktonic populations recovered from the same continuous culture. The second principal component (PC2) explained 20% of the total variance and generally separated samples by substrate. Populations growing on glucose tended to have the highest values of PC2, which were slightly higher than those of populations growing on cellobiose. Values of PC2 for both cellulose-adherent and cellulose-planktonic populations were lower than those of populations growing on soluble sugars, but were lowest at the lowest dilution rate. Figure 4.2 shows the dendrogram resulting from hierarchical clustering of the sample expression profiles. Two major clusters were identified, as well as one outlier which was a single replicate from the cellobiose condition at the lowest growth rate ($D \sim 0.02\ h^{-1}$). All other replicates

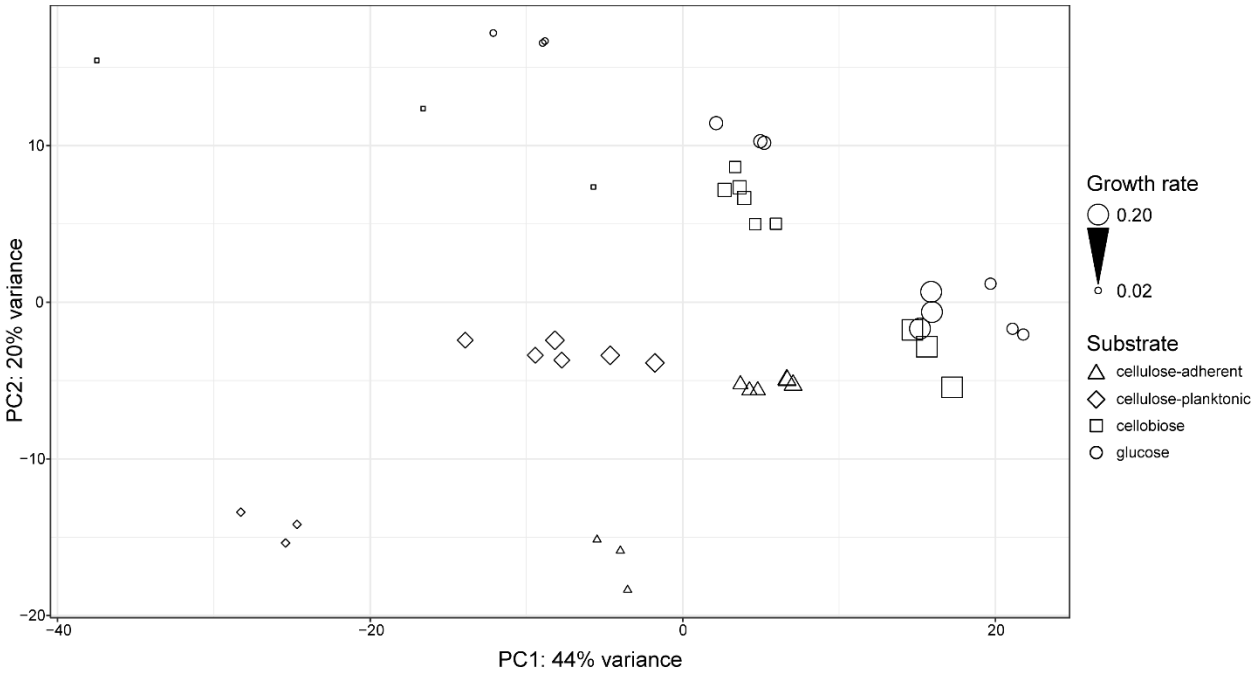


Figure 4.1: Principal components analysis of transcriptomes. The scatter plot shows the transcriptomes plotted according to the first two dimensions of principal components analysis. Cellulose-adherent samples are plotted as open triangles (Δ), cellulose-planktonic samples are plotted as open diamonds (◊), cellobiose samples are plotted as open squares (□), and glucose samples are plotted as open circles (○). The sizes of the shapes increase as a function of growth rate.

clustered together indicating good reproducibility within a condition at a given growth rate. The major clusters segregated by growth rate with the exception of the cellulose-planktonic samples which all clustered, regardless of growth rate, with only the slowest growing populations from the other conditions. Within the two major clusters, the samples grouped primarily according to substrate type.

General patterns of gene expression by *F. succinogenes* S85

A total of 3,119 genes were identified in the S85 transcriptome. Among these 3,119 expressed genes were: 718 (23.02%) genes significantly affected by growth rate (false discovery rate (FDR) < 0.01 and log2 fold change (LFC) > 1), and 806 (25.84%) genes significantly

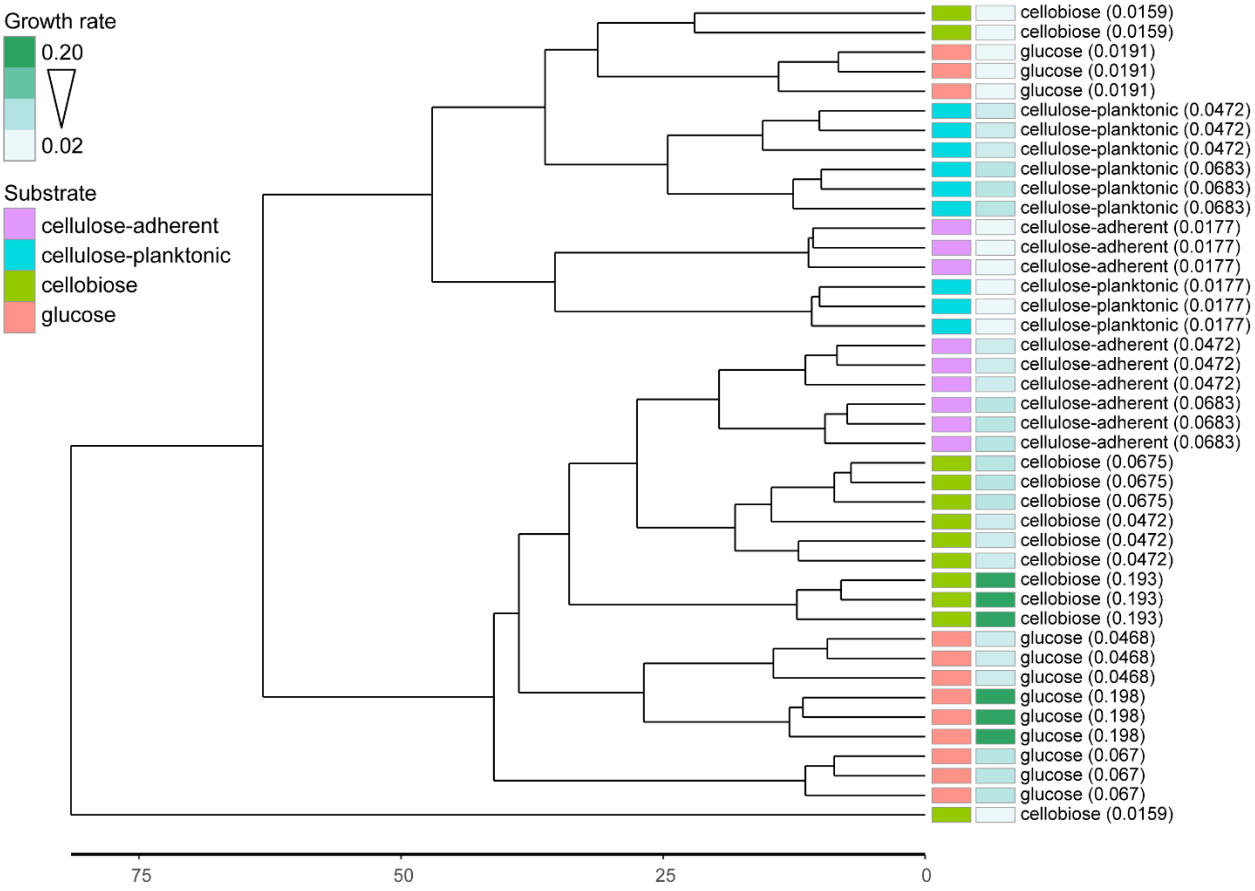


Figure 4.2: Hierarchical clustering of transcriptomes. The dendrogram shows the results of complete-linkage hierarchical clustering of the transcriptomes based on Euclidean distance. Samples are marked by colored according to substrate as follows: purple = cellulose-adherent, blue = cellulose-planktonic, green = cellobiose, red = glucose. Samples are separately colored according to growth rate from light (slow) to dark green (fast). Samples are labeled according to substrate type followed by empirically determined growth rates in parentheses.

affected by carbon substrate (FDR < 0.01 and LFC > 1). Evidence of an effect on expression by both growth rate and substrate was observed for 356 (11.41%) genes. Hierarchical clustering of all genes identified 14 distinct clusters of genes that grouped together according to similar expression profiles across the different growth rates and carbon substrates (Table 4.2). All but two of the 3,119 genes were placed into the 14 clusters. The two outliers exhibited the largest increases in expression during growth on cellulose relative to growth on glucose and cellobiose observed for any genes, and were identified as Fisuc_0384 and Fisuc_0385. The two largest

Table 4.2: Significant enrichments among clusters of co-regulated genes

Cluster	# of Genes	Growth Rate	Substrate	KEGG Category		CAZyme	
0	2						
1	1601	NS	NS	Aminoacyl-tRNA biosynthesis	9.75×10^{-6}	NS	
				Chromosome and associated proteins	2.99×10^{-2}		
				DNA replication proteins	4.45×10^{-2}		
				Oxidative phosphorylation	2.14×10^{-3}		
				Ribosome biogenesis	1.02×10^{-4}		
				Transfer RNA biogenesis	3.91×10^{-7}		
2	433	NS	NS	NS		NS	
3	264	9.65×10^{-5}	1.93×10^{-8}	NS		GH45	0.02
4	220	2.13×10^{-5}	1.72×10^{-22}	Fatty acid biosynthesis	5.01×10^{-5}	GT4	0.03
				Lipid biosynthesis proteins	1.17×10^{-2}		
				Ribosome	8.70×10^{-7}		
5	138	NS	NS	NS		NS	
6	82	4.35×10^{-3}	1.02×10^{-25}	Chaperones and folding catalysts	0.01	NS	
7	74	5.25×10^{-28}	3.66×10^{-36}	NS		NS	
8	58	NS	4.18×10^{-2}	NS		NS	
9	58	3.12×10^{-31}	6.51×10^{-9}	ABC transporters	2.98×10^{-4}	NS	
				Chloroalkane and chloroalkene degradation	3.73×10^{-9}		
				Cysteine and methionine metabolism	4.04×10^{-7}		
				Monobactam biosynthesis	4.86×10^{-2}		
				Nitrogen metabolism	4.04×10^{-7}		
				Selenocompound metabolism	2.71×10^{-4}		
				Sulfur metabolism	4.24×10^{-16}		
				Sulfur relay system	3.52×10^{-3}		
				Transporters	2.87×10^{-7}		
10	57	7.16×10^{-22}	1.38×10^{-5}	NS		CBM35	7.09×10^{-7}
						CE12	1.46×10^{-2}
						CE6	4.44×10^{-2}
11	47	5.90×10^{-28}	4.01×10^{-22}	Other glycan degradation	0.01	CBM6	1.24×10^{-34}
				Pentose and glucuronate interconversions	0.03	CE6	4.08×10^{-4}
				Sphingolipid metabolism	0.01	GH10	2.30×10^{-5}
						GH30	1.57×10^{-2}
						GH43	2.14×10^{-7}
						PL1	2.30×10^{-5}
12	38	1.07×10^{-4}	1.40×10^{-18}	Starch and sucrose metabolism	4.59×10^{-4}	GH5	2.38×10^{-3}
						GH9	0.03
13	27	NS	7.83×10^{-16}	Transcription machinery	0.01	NS	
14	20	NS	2.87×10^{-6}	NS		NS	

Values for Growth Rate, Substrate, KEGG Category, and CAZyme are p-values, after adjusting for multiple comparisons (hypergeometric test, Holm's correction (30))
NS = not significant

clusters resulting from hierarchical clustering of all genes, clusters 1 and 2, together represented 65.21% of the total genes expressed, containing 1601 genes and 433 genes, respectively. Cluster 1 was enriched for genes in the following KEGG categories: aminoacyl-tRNA biosynthesis, chromosome and associated protein, DNA replication proteins, oxidative phosphorylation, ribosome biogenesis, and transfer RNA biogenesis. Three genes predicted to be involved glycogen metabolism were also identified among the genes in cluster 1, consisting of two glycogen synthases (Fisuc_1515 and Fisuc_2067) and a glycogen phosphorylase (Fisuc_2097) that were highly expressed, independent of growth rate or substrate type. Glycogen turnover has been reported to be important in the overall carbon flux of *F. succinogenes* during growth on both cellulose and soluble sugars (14). No evidence for a systematic effect of growth rate or substrate type was observed for either cluster 1 or 2. The remaining 12 clusters contained a maximum of 264 and a minimum of 20 genes, and all but cluster 5 were enriched in genes affected by growth rate, substrate, or both growth rate and substrate (Table 4.2).

Genes affected by growth rate in *F. succinogenes* S85

Of the 718 genes affected by growth rate, 273 (38.02%) exhibited increased relative expression as growth rate increased and 445 (61.98%) showed increased relative expression as growth rate decreased (Table 4.3). Eight clusters were enriched for genes affected by growth rate: clusters 3, 4, 6, 7, 9, 10, 11, and 12 (Table 4.2). In cluster 9, 55 of the 58 genes (94.83%) showed increased relative expression as growth rate increased. Cluster 9 was also significantly enriched in genes annotated to the following KEGG categories: ABC transporters, chloroalkane and chloroalkene degradation, cysteine and methionine metabolism, monobactam biosynthesis, nitrogen metabolism, selenocompound metabolism, sulfur metabolism, sulfur relay system, and

Table 4.3: Total genes affected by growth rate and substrate

Pair-wise comparison		Total genes
Growth rate		718
Faster rate 273	Slower rate 445	718
Substrate		806
Cellulose (adherent) 159	Glucose 232	391
Cellulose (adherent) 120	Cellobiose 211	331
Cellulose (adherent) 1	Cellulose (planktonic) 308	309
Cellulose (planktonic) 357	Glucose 149	506
Cellulose (planktonic) 265	Cellobiose 106	371
Cellobiose 146	Glucose 135	281

Values are the number of genes overexpressed for a condition relative to the other condition in the pair-wise comparison (FDR < 0.01 & LFC > 1)

transporters (Table 4.2). Elevated expression of 79 of the 220 genes (35.91%) in cluster 4 was associated with higher rates of growth, while expression of only one gene in this cluster, Fisuc_2501, was associated with slower rates of growth. Cluster 4 was significantly enriched in the KEGG categories: fatty acid biosynthesis, lipid biosynthesis proteins, and ribosome; in addition to CAZyme family GT4 (Table 4.2). In cluster 11, 46 of the 47 genes (97.87%) exhibited increased relative expression during slower rates of growth, and this gene cluster was significantly enriched in the KEGG categories: other glycan degradation, pentose and glucuronate interconversions, and sphingolipid metabolism. Cluster 11 was also enriched in genes from the following CAZyme families: CBM6, CE6, GH10, GH30, GH43, and PL1. Cluster 10 contained a clear majority of genes, 48 of 57 (84.21%), displaying expressions associated with slower rates of growth, and this cluster was significantly enriched in CAZyme families: CBM35, CE12, and CE6. Several of the CAZyme families enriched in clusters 10 and

11 were also significantly enriched among all genes affected by growth rate (CBM35, CBM6, GH10, GH43, and PL1) compared to the S85 genome as a whole (Table 4.4).

Genes affected by substrate in *F. succinogenes* S85

Of the 806 total genes affected by substrate, 391 differed in expression between cellulose-adherent populations and those growing on glucose; 331 were differentially expressed between cellulose-adherent populations and those growing on cellobiose; 309 were differentially expressed between cellulose-adherent and cellulose-planktonic cells; 506 were differentially expressed between cellulose-planktonic cells and those growing on glucose; 371 were differentially expressed between cellulose-planktonic cells and those growing on cellobiose; 281 differed in expression between populations growing on cellobiose and glucose (Table 4.3). A statistical trend (*Padj* = 0.074, Holm's correction (30)) supporting enrichment of the KEGG category carbohydrate metabolism was observed for genes affected by carbon substrate (Table 4.4). Populations growing on cellobiose tended to have the highest expression, regularized-log2 transformed sequence count (rlog), of genes in this category (mean rlog = 8.18 ± 1.89), followed by cellulose-planktonic cells (mean rlog = 7.89 ± 1.75), those growing on glucose (mean rlog = 7.39 ± 1.92), and finally cellulose-adherent cells (mean rlog = 7.35 ± 1.50). Genes showing an effect of substrate on their expression were also enriched for the following CAZyme families: CBM35, CBM6, CE6, GH10, GH43, and PL1 (Table 4.4).

Evidence for significant enrichment in genes showing an effect of carbon substrate on expression was observed for 11 of the 14 total gene clusters: clusters 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, and 14 (Table 4.2). Genes in cluster 12 generally showed increased expression in cellulose-planktonic populations (mean rlog = 10.11 ± 2.88) and cellulose-adherent populations (mean rlog

Table 4.4: KEGG categories and CAZyme families affected by growth rate and substrate

Effect	KEGG category	P_{adj}*	CAZyme family	P_{adj}*
Growth rate	Cysteine and methionine metabolism	0.09	CBM35	1.51×10^{-4}
			CBM6	3.57×10^{-10}
			GH10	2.72×10^{-2}
			GH43	1.64×10^{-7}
			PL1	2.72×10^{-2}
Substrate	Carbohydrate metabolism	0.07	CBM35	7.89×10^{-7}
			CBM6	5.62×10^{-11}
			CE6	3.77×10^{-2}
			GH10	2.64×10^{-3}
			GH43	7.89×10^{-7}
			PL1	2.64×10^{-3}

*hypergeometric test, Holm's correction (30)

= 9.16 ± 2.45) versus those growing on cellobiose (mean rlog = 8.25 ± 2.58) or glucose (mean rlog = 7.74 ± 2.42). Cluster 12 was significantly enriched for genes in CAZyme families GH5 and GH9, known to contain cellulases (see below), as well as the KEGG category starch and sucrose metabolism. A high percentage (> 75%) of genes in clusters 6, 7, and 11 showed elevated expression in cellulose-planktonic cells relative to the cellulose-adherent population. KEGG categories enriched in these clusters include: chaperones and folding catalysts, other glycan degradation, pentose and glucuronate interconversions, and sphingolipid metabolism (Table 4.2). Although a total of 308 genes demonstrated elevated expression in cellulose-planktonic cells relative to cellulose-adherent cells, only one gene, Fisuc_1093, was identified with elevated expression in cellulose-adherent cells relative to the cellulose-planktonic population.

Expression of cellulases and hemicellulases by *F. succinogenes* S85

A total of 208 CAZymes, encoded by 171 genes, were identified in the S85 transcriptome. The CAZyme exhibiting the highest mean gene expression across all samples

was Fisuc_3111 (CBM11-GH51), the locus coding for Cel51A (also known as endoglucanase 2/endoglucanase F). In addition to being one of the most highly expressed genes in the transcriptome (> 99th percentile), Fisuc_3111 had a relatively low coefficient of variation (CV) for expression (2.23% across all samples), and was placed in cluster 2 by hierarchical clustering of all genes. Other highly expressed (> 90th percentile) and relatively stable (< 5% CV) CAZyme genes were: Fisuc_1802 (GH8, *cel8B*), Fisuc_0393 (GH9, *cel9H*), Fisuc_2900 (GH94, cellodextrin phosphorylase), Fisuc_1530 (GH18), Fisuc_2097 (GT35, glycogen phosphorylase), Fisuc_1224 (GH5), Fisuc_3049 (GH2), Fisuc_1219 (GH8), Fisuc_0668 (GH57), Fisuc_1932 (GH13), and Fisuc_2988 (GH23).

A clustered heatmap showing deviations from the mean rlog expression across the samples for genes (n = 32) annotated to the major CAZyme cellulose families in the S85 genome (GH5, GH8, GH9, GH45, and GH51) is shown in Figure 4.3. A significant effect of growth rate on expression was observed for 11 of the 32 cellulase genes (34.38%). For 9 of the 11 cellulases affected by growth rate, faster growth was associated with decreased expression regardless of substrate. A significant effect of substrate type on expression was observed for 16 of the 32 cellulase genes (50%). Nine of the 16 fell into two clusters at the very bottom of the cellulase heatmap (Figure 4.3). Cellulase genes affected by substrate generally exhibited the lowest levels of expression during growth on glucose. Two other distinct patterns of expression were observed among these cellulase genes. One pattern consisted of elevated expression during growth on cellulose, irrespective of attachment. The other was characterized by increased expression during growth on cellobiose as well as in cellulose-adherent and cellulose-planktonic cells. Hierarchical clustering of all S85 genes based on similar gene expression patterns identified cluster 12 as being enriched for cellulase families GH5 and GH9 (Table 4.2).

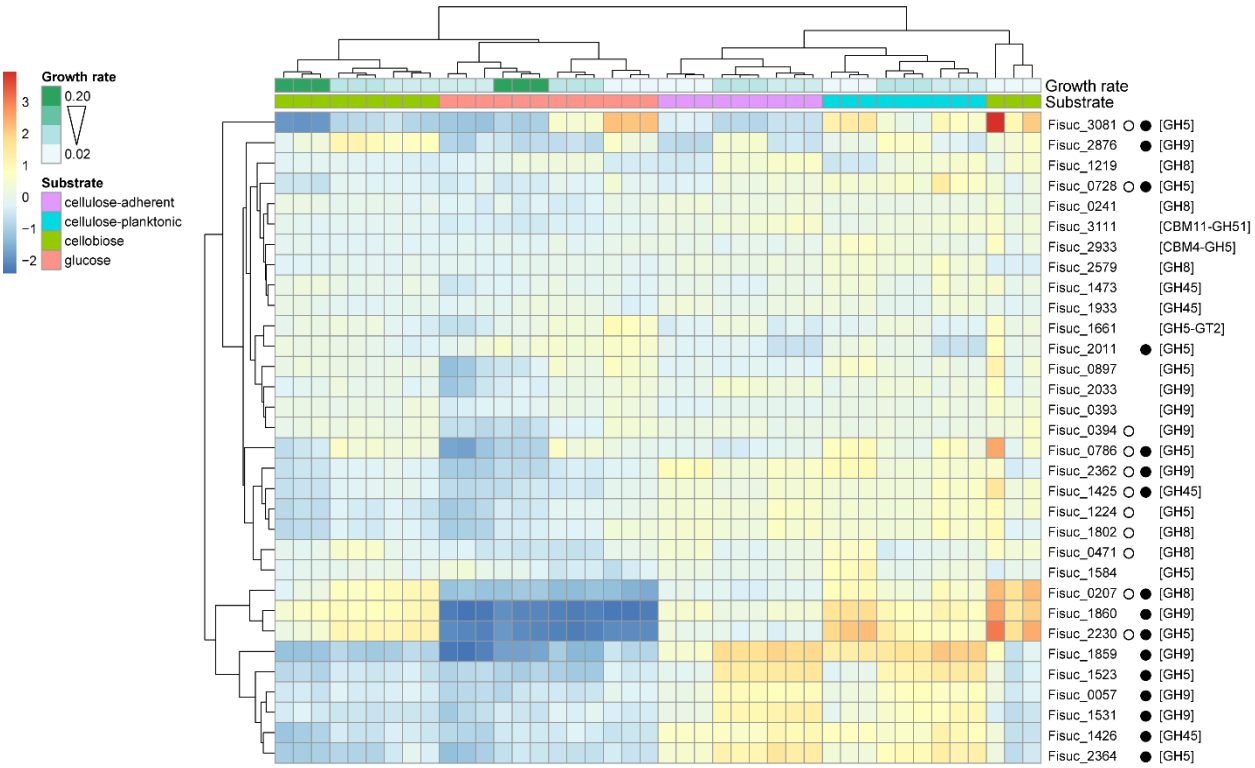


Figure 4.3: Expression of cellulases. The heatmap shows the deviation from the mean rlog expression across samples for cellulase genes in CAZyme families: GH5, GH8, GH9, GH45, and GH51. Genes are arranged vertically according to the results of hierarchical clustering. Samples are arranged horizontally according to the results of hierarchical clustering and are marked by colored according to substrate (purple = cellulose-adherent, blue = cellulose-planktonic, green = cellobiose, red = glucose) and growth rate (slow = light green and fast = dark green). The heatmap is colored according to log2 fold change from the mean for gene in a given sample, with the intensity of red (positive change) or blue (negative change) indicating the magnitude of the effect. Open circles (○) to the right of the locus ID indicate an effect of growth rate on gene expression, while solid circles (●) indicated an effect of carbon substrate on expression. CAZyme families for the genes are in brackets.

A clustered heatmap showing deviations from the mean rlog expression across the samples for genes (n = 53) annotated to the major hemicellulase-related carbohydrate-binding module and glycoside hydrolase CAZyme families in the S85 genome (CBM6, CBM35, GH2, GH3, GH10, GH11, GH18, GH26, GH43) is shown in Figure 4.4. Three major patterns of gene expression were identified in the hemicellulase heatmap (Figure 4.4), two of which varied considerably with substrate type and growth rate. One of the two variable expression patterns,

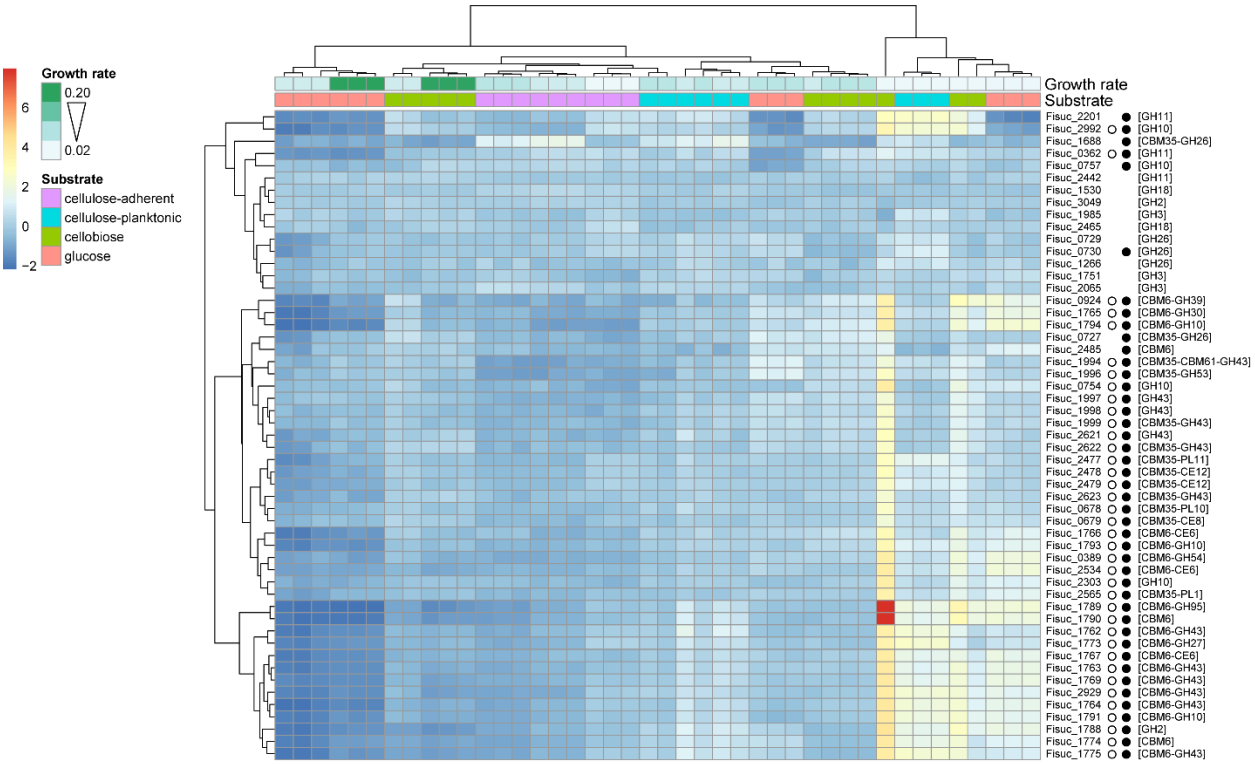


Figure 4.4: Expression of hemicellulases. The heatmap shows the deviation from the mean rlog expression across samples for hemicellulase genes in CAZyme families: CBM6, CBM35, GH2, GH3, GH10, GH11, GH18, GH26, and GH43. Genes are arranged vertically according to the results of hierarchical clustering. Samples are arranged horizontally according to the results of hierarchical clustering and are marked by colored according to substrate (purple = cellulose-adherent, blue = cellulose-planktonic, green = cellobiose, red = glucose) and growth rate (slow = light green and fast = dark green). The heatmap is colored according to log2 fold change from the mean for gene in a given sample, with the intensity of red (positive change) or blue (negative change) indicating the magnitude of the effect. Open circles (○) to the right of the locus ID indicate an effect of growth rate on gene expression, while solid circles (●) indicated an effect of carbon substrate on expression. CAZyme families for the genes are in brackets.

consisting of elevated expression at the slowest growth rate (D ~ 0.02 h^{-1}) in glucose, cellobiose, and cellulose-planktonic populations, was demonstrated primarily by hemicellulases containing a carbohydrate-binding module family 6 (CBM6) domain. The other was characterized by elevated expression at the slowest growth rate (D ~ 0.02 h^{-1}) during growth on cellobiose and occasionally glucose, but not cellulose-planktonic growth. Most of the hemicellulase genes that possess a CBM35 domain exhibited this pattern of expression, as did the remainder of those

encoding enzymes with a CBM6 domain. Overall, a significant effect of growth rate on expression was detected for 38 of the 53 hemicellulases (71.70%). The expression of hemicellulase genes generally decreased as growth rate increased, regardless of substrate. Four CAZyme families implicated in hemicellulase breakdown by *Fibrobacter* (CBM35, CBM6, GH10, and GH43) were significantly enriched among the complete set of S85 genes determined to be affected by growth rate (Table 4.4). Among the hemicellulase-related CAZymes, a clear majority, 44 of 53 (83.02%), exhibited evidence of a significant effect of substrate type on their expression. Most demonstrated elevated expression in cells growing on soluble sugars relative to cellulose-adherent cells, 30/44 for glucose versus adherent, 40/44 for cellobiose versus adherent, and 29/44 for cellulose-planktonic versus adherent. Additionally, 29 of the 44 were more highly expressed during growth on cellobiose relative to growth on glucose. CBM35, CBM6, CE6, GH10, GH43, and PL1, CAZyme families implicated in the breakdown of hemicellulose by *Fibrobacter*, were also significantly enriched among all S85 genes affected by substrate type (Table 4.4). Hierarchical clustering of the expression profiles of all genes identified two clusters that were significantly enriched in hemicellulase-related CAZyme familes (Table 4.2). Cluster 10 was enriched in genes annotated to CBM35 and GH43, as well as PL1, while cluster 11 was enriched in the families CBM6, GH10, and GH43, in addition to CE6, GH30, and PL1.

4.4 Discussion

In this study we examined gene expression by *F. succinogenes* S85 at steady-state in continuous culture on several carbon substrates over a range of different growth rates. Our data suggest that growth rate had a larger overall effect on the total S85 transcriptome than did any particular carbon substrate. This is supported by the results of PCA comparing the overall

transcriptomes as well as hierarchical clustering of the samples. A total of 718 genes were identified as being affected by growth rate. Most of these 718 (61.98%) showed increased relative expression during slower rates of growth compared to cultures growing at a faster rate, after controlling for the effect of substrate. Many of these genes are annotated as being involved in glycan degradation and are elaborated upon below. These results find some analogy to gene expression patterns in *Clostridium thermocellum* (phylum Firmicutes), one of the most intensively studied cellulolytic bacteria. In *C. thermocellum*, the total number of genes affected by growth rate exceeded the number affected by substrate type (31). Moreover, two of the major genes associated with cellulose degradation, namely *cipA*, the cellulosomal scaffoldin, and *celS*, the major exoglucanase, display markedly higher expression at lower growth rates in cellobiose-limited cultures (32).

Of the genes that showed the opposite pattern of expression (*i.e.*, increased relative expression at faster growth rates), many are annotated to functional categories related to sulfur metabolism. It is not clear why genes for sulfur metabolism should be induced under any of the culture conditions examined in this study. The use of cysteine as a reducing agent in the media, which is readily utilized as a source of sulfur by *F. succinogenes*, should ensure that the concentration of sulfur always exceeds cellular requirements.

A large group of genes with relatively elevated expression in strain S85 during slow growth are annotated to CAZyme families believed to be responsible for hemicellulose hydrolysis by *Fibrobacter*. These same CAZymes were also among those whose expression was most affected by carbon substrate. A notable characteristic of the Fibrobacteres is their inability to utilize the constituents of hemicellulose (other than glucose) for growth, despite readily hydrolyzing this class of polysaccharides (33, 34). This observation has led to the hypothesis

that these bacteria utilize these enzymes only as a means of gaining access to cellulose, which is typically embedded in a matrix of hemicellulose, pectin, and lignin (21). Tight control over the genes coding for *Fibrobacter* hemicellulases is consistent with this hypothesis because these enzymes are apparently only required under very specific circumstances. A slow growth rate may signal to the cell that the preferred substrate, cellulose, is not readily available and that the production of hemicellulases may assist in gaining access to it. Our results also suggest that *Fibrobacter* cells are able to distinguish between slow rates of growth while attached to cellulose versus slow growth on glucose or cellobiose because cellulose-adherent was the only substrate condition examined where genes for hemicellulases did not show elevated expression at the slowest growth rate examined ($D \sim 0.02\ h^{-1}$). One hypothesis is that the presence of glucose or cellobiose in combination with a low rate of growth signals to the cell that cellulose is near, but not readily available. Under these conditions the induction of genes for hemicellulases would be beneficial. Moreover, the particular sensitivity of the *Fibrobacter* cellulase system to feedback inhibition (35) potentially means that, even during relatively slow growth on cellulose, cellobiose and glucose are consumed immediately as they are produced, ensuring that the signal for hemicellulase induction remains low.

It is well established that cellulose degradation by *Fibrobacter* likely involves many enzymes, working synergistically, but the conditions under which they are expressed are less well understood. Our observations support a model with several distinct patterns of expression among cellulase genes by strain S85. The CAZyme gene showing the highest levels of expression across growth rates and carbon substrates was Fisuc_3111, the gene encoding Cel51A (also known as endoglucanase 2/endoglucanase F). This enzyme was one of the first endoglucanases to be purified from *F. succinogenes*, in addition to being one of the first major

cellulose-binding proteins identified (36, 37). Cel51A is also the only GH51 family protein in the S85 genome (21). Our results suggest that this enzyme is expressed constitutively, based on a low coefficient of variation in its expression across all conditions. This is consistent with proteomic data which identified Cel51A in the outer membrane of S85 regardless of whether it was cultured on glucose, acid-swollen cellulose, or microcrystalline cellulose (38). Additionally, Cel51A has been implicated in outer membrane blebbing and vesicle formation because the protein is largely cell-associated until encountering cellulose at which point it becomes concentrated in areas of blebbing and membrane disruption as it translocates to the cellulose surface, taking pieces of the outer membrane with it (37, 39). Constitutive expression of the gene for Cel51A would ensure that plenty of this cellulase is available even as it is torn out of the membrane upon contact with cellulose. Along with Fisuc_3111, our data suggest that 12 of the 32 (37.50%) cellulase genes in the S85 genome are constitutively expressed by this organism. These genes represent all families of cellulases possessed by S85: GH5, GH8, GH9, GH45, and GH51. Constitutive expression of at least a subset of cellulases explains why *Fibrobacter* does not exhibit a lag in growth when transferred to cellulose after being grown on soluble sugars (28).

Two additional patterns of expression were observed among the cellulase genes of S85. One pattern, represented by Fisuc_0207, Fisuc_1860 (encoding Cel9C/CelD), and Fisuc_2230 (encoding Cel5G/Cel3), consisted of very low relative expression on glucose and increased expression in both cellulose subpopulations as well as in cells growing on cellobiose. These three cellulase genes were most highly expressed in cellulose-planktonic and cellobiose populations, particularly at the slowest growth rate. These results suggest that cellobiose, either when fed directly or produced from cellulose hydrolysis, is involved in stimulating the

expression of these genes. Another major pattern of expression observed among the S85 cellulases included at least six genes, one of which was Fisuc_1859, the gene for Cel9B (also known as endoglucanase 1/endoglucanase E). These cellulase genes displayed elevated expression in cellulose-adherent and cellulose-planktonic populations, but not during growth on cellobiose. Moreover, their expression tended to increase along with growth rate during growth on cellulose, which was atypical for glycoside hydrolases generally in this study and also in studies of gene expression in *C. thermocellum* (31, 32). Cel9B, endoglucanase 1, has been shown to be released during growth by S85 and to contribute a large fraction of the extracellular endoglucanase activity produced by this organism (36, 39). Our results confirm previous reports that expression of this enzyme is regulated by substrate, namely repressed by cellobiose, which is also the major product produced by this enzyme (39). Tight control over the expression of cel9B likely helps to ensure that it is efficiently delivered to its substrate and that the substrate is already in close proximity to the cell so that the cellobiose produced can be rapidly consumed before it is lost to competitors.

In addition to hydrolytic enzymes, proteins that promote cellular attachment to cellulose are critically important for *Fibrobacter* because adherence appears to be essential for efficient growth of these bacteria on cellulose (18). *Fibrobacter* proteins containing a fibro-slime domain have been implicated in attachment to cellulose in these bacteria (22). The S85 genome has ten copies of genes predicted to code for fibro-slime proteins (21). Our data show that several of these are among the most highly expressed genes in the S85 transcriptome, with three in the top 98th percentile across all conditions: Fisuc_1979, Fisuc_1474, and Fisuc_0377. Previously, we reported that 8 of the 10 fibro-slime genes were upregulated during growth on cellulose relative to growth on glucose (40). However, the data generated in this study suggests that a majority of

fibro-slime genes are expressed constitutively. Evidence of an effect on expression by growth rate and carbon source was observed for only two and three fibro-slime genes, respectively. The most likely explanation for the discrepancy is that, because the previous work examined populations undergoing log-phase growth in batch cultures, differences in growth rate between the substrates confounded the interpretation of the differences in gene expression. These results emphasize the importance of controlling for growth rate when evaluating gene expression, as shown previously for *C. thermocellum* (31, 32).

The *Fibrobacter* population in cellulose-limited continuous cultures is heterogeneous, consisting primarily of cells that are adherent to the cellulose surface as well as a much smaller fraction of cells that are planktonic (*i.e.*, not adherent). Microscopic observations of planktonic cells undergoing cell division support the conclusion that both subpopulations are engaging in cellular growth, but likely utilizing different substrates (41). As has been reported previously, we observed a low concentration of soluble sugar (~1 mM), consisting primarily of glucose, present in steady-state cellulose cultures, further demonstrating that substrate is available to feed planktonic cells under these conditions (42). Therefore, we fractionated these two subpopulations prior to RNA extraction in order to investigate their gene expression separately. The results identified clear differences in gene expression between them. Patterns of gene expression characteristic of cellulose-planktonic populations exhibited some commonalities with both cellulose-adherent populations and those growing on soluble sugars. Several cellulase genes displayed elevated expression in both cellulose-adherent and cellulose-planktonic populations, but not in populations growing on either soluble sugar. However, genes for hemicellulases, which were not induced in cellulose-adherent populations, showed elevated expression during slow growth in cellulose-planktonic populations as well as during slow growth

on glucose and cellobiose. A likely explanation for these observations is that cellulose-planktonic populations are heterogenous in nature, consisting of both the recently released daughter cells of the cellulose-adherent population, exhibiting gene expression patterns still resembling those of their mother cells, as well as cells that have undergone several cycles of division growing on the soluble sugars available in the liquid phase of the culture.

The observation that continuous cultures on all three substrates (cellulose, glucose and cellobiose) contained ~1 mM (glucose equivalent) of residual soluble sugars is consistent with the relatively poor affinity of this strain for soluble sugars (Ks ~ 0.5 mM for both glucose and cellobiose (43)). The average degree of polymerization of these sugars (based on a ratio of total sugars/reducing sugars) was ~1.4 across the three substrates fed, indicating that glucose was the most abundant of the residual sugars. The identities of the other sugars could include cellobiose (either remaining from cellobiose feeding, or produced by reversal of cellobiose phosphorylase), and small amounts of longer cellodextrins produced in the same manner (41). These residual sugars could also potentially include maltose and maltodextrins produced from turnover of glycogen, as these carbohydrates have been detected in extracellular media during batch culture growth of *F. succinogenes* S85 on soluble sugars (14). The intriguing potential role of glycogen turnover in *F. succinogenes* in cellular metabolism is supported by our observations that two annotated glycogen synthases (Fisuc_1515 and Fisuc_2067) and a glycogen phosphorylase (Fisuc_2097) were highly expressed, independent of growth rate or substrate type. Further research is needed to elucidate the interplay between the catabolism of cellulose and its hydrolytic products on the one hand, and glycogen synthesis and turnover on the other. Regardless, the fact that *F. succinogenes* displayed differential expression of nearly 300 genes

when grown on glucose versus cellobiose suggests that these two sugars have distinct roles in metabolic regulation of this bacterium.

This study provides the most detailed view of the *F. succinogenes* S85 transcriptome to date, and is the first systematic assessment of the effect of growth rate and carbon substrate on gene in expression in this important succinate-producing rumen bacterium. We have shown for the first time tight control over global expression of genes encoding hemicellulases in this bacteria as well as multiple distinct patterns of expression for cellulase genes. This work reinforces the importance of controlling for growth rate when evaluating effects on gene expression. We have also demonstrated differences in gene expression between adherent and planktonic populations in a bacterial culture growing on an insoluble substrate, a variable that is often overlooked in studies of anaerobic fiber-degraders. Future work focused on how the patterns of gene expression described here relate to the protein levels in *Fibrobacter* is warranted.

4.5 Materials & Methods

Bacterial strain, culture conditions and sampling

The *Fibrobacter succinogenes* type strain, S85 (ATCC 19169), was used for all experiments in this study. The medium used for cultivation was a modified version of the formula first described by Scott and Dehority (44), the exact contents of which are listed in Table 4.5. Carbon-limited continuous cultures were conducted at 39° C (pH = 6.8) using a bioreactor system capable of regulated delivery of an insoluble substrate as a segmented slurry, described previously (45). Carbon sources tested included glucose, cellobiose, and microcrystalline cellulose (Sigmacell 20, Sigma-Aldrich, St. Louis, MO). The soluble sugars were included in

Table 4.5: Basal media composition for continuous cultures

component	Final [mM]	Final [uM]
K_2HPO_4	6.84	-
Na_2CO_3	37.7	-
NaCl	15.4	-
NH_4Cl	13.5	-
$MgCl_2$	0.418	-
$CaCl_2$	0.449	-
$FeSO_4$	0.0719	-
$MnCl_2$	0.138	-
$ZnCl_2$	0.0138	-
$CoCl_2$	0.00812	-
Isobutyric acid	0.760	-
Isovaleric acid	0.656	-
2-methylbutyric acid	0.656	-
Valeric acid	0.656	-
Nicotinamide	-	1.64
Pyridoxine HCl	-	0.973
Thiamine HCl	-	0.593
Riboflavin	-	0.531
Ca-D-pantothenate	-	0.42
p-Aminobenzoic acid	-	0.0729
Biotin	-	0.0205
Folic acid	-	0.00283
Tetrahydrofolic acid	-	0.00281
Vitamin B12	-	0.00148
resazurin	0.00796	-
L-Cysteine-HCl	2.85	-

gas phase = 100% CO_2

the medium at 2 g L^{-1}, while the concentration of cellulose used was 3.6 g L^{-1}. The higher concentration of cellulose was used because, unlike for the soluble sugars, a substantial fraction of the cellulose is not consumed during continuous culture of *F. succinogenes* S85 (42). The amount of cellulose used was determined from previous empirical observations with the goal of normalizing overall substrate consumption, cell growth, and product formation. Sampling of continuous cultures growing on glucose or cellobiose was performed at four targeted dilution rates: D = 0.018, 0.047, 0.068, and 0.20 h^{-1}. Sampling of continuous cultures growing on

crystalline cellulose was performed at three targeted dilution rates: D = 0.018, 0.047, and 0.068 h^{-1}. Samples corresponding with the fastest dilution rate evaluated for the soluble sugars (D = 0.20 h^{-1}) could not be collected from cellulose-limited continuous cultures because wash-out occurs for this organism growing on crystalline cellulose at dilution rates greater than 0.08 h^{-1} (42). Three samples were collected at 5 - 20 h intervals for each carbon source and dilution rate combination at steady state (> 3 reactor volumes, equivalent to > 97% volume turnover) for transcriptomic analysis as well as quantification of fermentation products, residual soluble sugars and, in the case of the cellulose-limited continuous cultures, residual cellulose in the reactor. Consistent with chemostat theory, growth rate was assumed to equal the dilution rate for steady-state continuous cultures.

Substrate and product analysis

Supernatants collected from continuous cultures at steady state were centrifuged for 5 min at 12,000 x g and analyzed for total soluble sugar content using a phenol/sulfuric acid method (46), with glucose as the standard. Reducing sugars were measured by the dinitrosalicylic acid method, using glucose as the standard (47). Supernatants were also analyzed for fermentation acids and alcohols by high performance liquid chromatography (HPLC) after treatment with calcium hydroxide and cupric sulfate, with identical chromatographic conditions to those reported previously (45). Two technical replicates were analyzed for each biological replicate, and the results averaged. Normality and homoscedasticity of the data were assessed and linear regression and analysis of variance (ANOVA) performed to determine the effect of growth rate, carbon source, and potential interaction between the two explanatory variables on the production of the major fermentation products, succinate and

acetate. Residual cellulose was determined gravimetrically after treatment with boiling neutral detergent (48) to remove adherent bacterial cells, and subtracted from the initial concentration to calculate the grams of cellulose consumed per liter. Mean values are reported with accompanying standard deviations.

RNA extraction, library prep and sequencing

Total RNA was extracted as described previously (49). Cells from 10 mL of glucose-limited or cellobiose-limited continuous cultures were harvested via centrifugation (12,000 x g, 10 min, 5° C) and resuspended in 1 mL RNA extraction buffer (50 mM sodium acetate, 10 mM EDTA, pH = 5.1). Cells from 10 mL of cellulose-limited continuous cultures were first separated into cellulose-adherent and cellulose-planktonic fractions by low speed centrifugation at 1,000 x g for 5 min. The supernatant, containing the cellulose-planktonic population, was then collected and processed separately. Cells from both the cellulose-adherent and cellulose-planktonic populations were subsequently recovered by centrifugation (12,000 x g, 10 min, 5° C) and resuspended in 1 mL RNA extraction buffer, as above. The cell suspension was combined with 0.5 g of 0.1 mm zirconium beads, 700 µL equilibrated phenol (pH < 5.0) and 50 µL 20% SDS. The sample was subjected to bead-beating for 2 min, followed by incubation at 60° C for 10 min, followed by additional bead-beating for 2 min. The organic and aqueous phases were separated via centrifugation and the aqueous phase extracted with 500 µL equilibrated phenol. Four additional extractions were performed, two with phenol:chloroform (50:50) followed by two with chloroform only. The RNA was precipitated with 3 M Na acetate and isopropanol, and resuspended in RNase-free H_2O. Residual DNA contamination was removed using TURBO DNA-free (Life technologies, Carlsbad, CA). Total RNA was quantified using the RNA HS

assay kit and Qubit® Fluorometer (Invitrogen, Carlsbad, CA) prior to shipment to the DOE Joint Genome Institute (Walnut Creek, CA) for library preparation and sequencing. rRNA was removed from 100 ng of total RNA using Ribo-Zero™ rRNA Removal Kit (Epicentre, Madison, WI). Stranded cDNA libraries were generated using the Illumina Truseq Stranded RNA LT kit (Illumina, San Diego, CA). The rRNA depleted RNA was fragmented and reversed transcribed using random hexamers and SSII (Invitrogen) followed by second strand synthesis. The fragmented cDNA was treated with end-pair, A-tailing, adapter ligation, and 10 cycles of PCR. qPCR was used to determine the concentration of the libraries. Libraries were sequenced on the Illumina Hiseq-2500 using the 2 x 100 nt run mode.

Preprocessing, alignment and counting of sequencing reads

Raw reads were filtered and trimmed using BBDuk (https://sourceforge.net/projects/bbmap/). Specifically, raw reads were evaluated for artifact sequence by kmer matching (kmer=25), allowing 1 mismatch, and detected artifacts trimmed from the 3' end of the reads. RNA spike-in reads, PhiX reads and reads containing any Ns were removed. Quality trimming was performed using the phred trimming method set at Q10, and any reads under 45 bases in length were removed. The resulting reads from each library were aligned to the *F. succinogenes* S85 reference genome (IMG Taxonoid 646311927) using BBMap (https://sourceforge.net/projects/bbmap/) with only unique mappings allowed. If a read mapped to more than one location it was discarded. featureCounts (50) was used to generate raw gene counts. The extent of correlation between biological samples was evaluated from the raw gene counts using Pearson's correlation.

Differential expression

Raw gene counts were imported into R (https://www.r-project.org/) and tested for differential expression using the "DESeq2" package (http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html) (51). A DESeqDataSet object was generated for differential expression analysis from the complete data set with the design formula: expression ~ growth rate + substrate type, with growth rate treated as a continuous variable and substrate type treated as a categorical variable. Empirically determined dilution rates were used as values for growth rate. The likelihood ratio test was used to determine which genes showed evidence of a statistically significant effect of either carbon source or growth rate on expression by comparing the full model, containing both explanatory variables, to a reduced model without the variable of interest. Independent hypothesis weighting was performed using the R Bioconductor package "IHW" (52) in order to optimize the power of the statistical tests. The false discovery rate (FDR) was controlled at < 0.01 using the method of Benjamini and Hochberg (53). A log2 fold-change (LFC) > 1 was used as the criterion for biological significance.

Ordination and clustering

Raw gene counts were transformed using the regularized-logarithm transformation (rlog), log base 2, prior to exploratory analyses to ensure that the gene expression data were normalized across samples and approximately homoscedastic (51). Mean rlog values are presented along with ± 1 standard deviation. Principal components analysis (PCA) of sample-to-sample distances was performed by passing the rlog-transformed data to the *plotPCA* function of "DESeq2" (51). Clustering of samples was performed by calculating the Euclidean distance

between the samples from the rlog-transformed data and passing the resulting distance matrix to the *hclust* function from base R, which by default performs complete-linkage hierarchical clustering. Clustering of genes was performed by first calculating the deviation from the mean in each sample for all genes by subtracting a gene's average expression across all samples from its expression in a single sample. Euclidean distances between the genes were then calculated from the deviations and used as input for clustering using *hclust*. Gene clusters were extracted from the dendrogram using the *cutreeDynamic* function from the R package "dynamicTreeCut" (54) with a minimum cluster size of 20. Clustered heatmaps of deviations for select genes were generated using *pheatmap* from the R package "pheatmap".

Gene annotation

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations were determined using the BlastKOALA online annotation tool (http://www.kegg.jp/blastkoala/) for K number assignment (55). Carbohydrate-active enzyme (CAZyme) annotation was performed using HMMER (56) and the dbCAN database (57). CAZyme annotations with e-value > 1×10^{-18} and coverage < 0.35 were discarded, as recommended for bacteria (http://csbl.bmb.uga.edu/dbCAN/). The hypergeometric test was used to determine significant enrichment of a functional category among a particular subset of genes of interest. P-values were adjusted to control the family-wise error rate across multiple comparisons using the base R function *p.adjust*, which uses Holm's correction method by default (30).

Sequence accession numbers

Raw Illumina sequencing reads have been deposited in the National Center for Biotechnology Information's Short Read Archive (SRA) database under accessions PRJNA402842 - PRJNA402853 and PRJNA402860 - PRJNA402889.

4.6 References

1. Cosgrove DJ. 2005. Growth of the plant cell wall. Nat Rev Mol Cell Biol 6:850–861.
2. O'Sullivan AC. 1997. Cellulose: the structure slowly unravels. Cellulose 4:173–207.
3. Loqué D, Scheller HV, Pauly M. 2015. Engineering of plant cell walls for enhanced biofuel production. Current Opinion in Plant Biology 25:151–161.
4. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577.
5. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol 6:121–131.
6. Henderson G, Cox F, Ganesh S, Jonker A, Young W, Collaborators GRC, Abecia L, Angarita E, Aravena P, Arenas GN, Ariza C, Attwood GT, Avila JM, Avila-Stagno J, Bannink A, Barahona R, Batistotti M, Bertelsen MF, Brown-Kav A, Carvajal AM, Cersosimo L, Chaves AV, Church J, Clipson N, Cobos-Peralta MA, Cookson AL, Cravero S, Carballo OC, Crosley K, Cruz G, Cucchi MC, Barra R de la, Menezes ABD, Detmann E, Dieho K, Dijkstra J, Reis WLS dos, Dugan MER, Ebrahimi SH, Eythórsdóttir E, Fon FN, Fraga M, Franco F, Friedeman C, Fukuma N, Gagić D, Gangnat I, Grilli DJ, Guan LL, Miri VH, Hernandez-Sanabria E, Gomez AXI, Isah OA, Ishaq S, Jami E, Jelincic J, Kantanen J, Kelly WJ, Kim S-H, Klieve A, Kobayashi Y, Koike S, Kopecny J, Kristensen TN, Krizsan SJ, LaChance H, Lachman M, Lamberson WR, Lambie S, Lassen J, Leahy SC, Lee S-S, Leiber F, Lewis E, Lin B, Lira R, Lund P, Macipe E, Mamuad LL, Mantovani HC, Marcoppido GA, Márquez C, Martin C, Martinez G, Martinez ME, Mayorga OL, McAllister TA, McSweeney C, Mestre L, Minnee E, Mitsumori M, Mizrahi I, Molina I, Muenger A, Muñoz C, Murovec B, Newbold J, Nsereko V, O'Donovan M, Okunade S, O'Neill B, Ospina S, Ouwerkerk D, Parra D, Pereira LGR, Pinares-Patiño C, Pope PB, Poulsen M, Rodehutscord M, Rodriguez T, Saito K, Sales F, Sauer C, Shingfield K, Shoji N, Simunek J, Stojanović-Radić Z, Stres B, Sun X, Swartz J, Tan ZL, Tapio I, Taxis TM, Tomkins N, Ungerfeld E, Valizadeh R, Adrichem P van, Hamme JV, Hoven WV, Waghorn G, Wallace RJ, Wang M, Waters SM, Keogh K, Witzig M, Wright A-DG, Yamano H, Yan T, Yáñez-Ruiz DR, Yeoman CJ, Zambrano R, Zeitz J, Zhou M, Zhou HW, Zou CX, Zunino P, 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.
7. Weimer PJ. 1992. Cellulose degradation by ruminal microorganisms. Critical Reviews in Biotechnology 12:189–223.
8. Kobayashi Y, Shinkai T, Koike S. 2008. Ecological and physiological characterization shows that *Fibrobacter succinogenes* is important in rumen fiber digestion - review. Folia Microbiol (Praha) 53:195–200.
9. Halliwell G, Bryant MP. 1963. The cellulolytic activity of pure strains of bacteria from the rumen of cattle. J Gen Microbiol 32:441–448.
10. Dehority BA, Scott HW. 1967. Extent of cellulose and hemicellulose digestion in various forages by pure cultures of rumen bacteria. J Dairy Sci 50:1136–1141.
11. Bryant MP, Doetsch RN. 1954. A study of actively cellulolytic rod-shaped bacteria of the bovine rumen. J Dairy Sci 37:1176–1183.

12. Jun HS, Qi M, Ha JK, Forsberg CW. 2007. *Fibrobacter succinogenes*, a dominant fibrolytic ruminal bacterium: Transition to the post genomic era. Asian-Australasian Journal of Animal Sciences, Asian-Australasian Journal of Animal Sciences 20:802–810.
13. Maglione G, Russell JB, Wilson DB. 1997. Kinetics of cellulose digestion by *Fibrobacter succinogenes* S85. Appl Environ Microbiol 63:665–669.
14. Nouaille R, Matulova M, Pätoprstý V, Delort A-M, Forano E. 2009. Production of oligosaccharides and cellobionic acid by *Fibrobacter succinogenes* S85 growing on sugars, cellulose and wheat straw. Appl Microbiol Biotechnol 83:425–433.
15. Li Q, Siles JA, Thompson IP. 2010. Succinic acid production from orange peel and wheat straw by batch fermentations of *Fibrobacter succinogenes* S85. Appl Microbiol Biotechnol 88:671–678.
16. Gokarn RR, Eiteman MA, Martin SA, Eriksson KE. 1997. Production of succinate from glucose, cellobiose, and various cellulosic materials by the ruminal anaerobic bacteria *Fibrobacter succinogenes* and *Ruminococcus flavefaciens*. Appl Biochem Biotechnol 68:69–80.
17. Gong J, Forsberg CW. 1989. Factors affecting adhesion of *Fibrobacter succinogenes* subsp. *succinogenes* S85 and adherence-defective mutants to cellulose. Appl Environ Microbiol 55:3039–3044.
18. Kudo H, Cheng KJ, Costerton JW. 1987. Electron microscopic study of the methylcellulose-mediated detachment of cellulolytic rumen bacteria from cellulose fibers. Can J Microbiol 33:267–272.
19. Miron J, Ben-Ghedalia D, Morrison M. 2001. Invited review: adhesion mechanisms of rumen cellulolytic bacteria. J Dairy Sci 84:1294–1309.
20. Leschine SB. 1995. Cellulose degradation in anaerobic environments. Annu Rev Microbiol 49:399–426.
21. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D, Brumm PJ. 2011. The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. PLoS ONE 6:e18814.
22. Jun H-S, Qi M, Gong J, Egbosimba EE, Forsberg CW. 2007. Outer membrane proteins of *Fibrobacter succinogenes* with potential roles in adhesion to cellulose and in cellulose digestion. J Bacteriol 189:6806–6815.
23. Arntzen MØ, Várnai A, Mackie RI, Eijsink VGH, Pope PB. 2017. Outer membrane vesicles from *Fibrobacter succinogenes* S85 contain an array of carbohydrate-active enzymes with versatile polysaccharide-degrading capacity. Environ Microbiol 19:2701–2714.
24. Forsberg CW, Beveridge TJ, Hellstrom A. 1981. Cellulase and xylanase release from *Bacteroides succinogenes* and its importance in the rumen environment. Appl Environ Microbiol 42:886–896.
25. Brumm P, Mead D, Boyum J, Drinkwater C, Deneke J, Gowda K, Stevenson D, Weimer P. 2011. Functional annotation of *Fibrobacter succinogenes* S85 carbohydrate active enzymes. Appl Biochem Biotechnol 163:649–657.
26. Malburg Jr. LM, Forsberg CW. 1993. *Fibrobacter succinogenes* S85 possesses at least nine distinct glucanase genes. Can J Microbiol 39:882–891.
27. Qi M, Jun H-S, Forsberg CW. 2007. Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. Appl Environ Microbiol 73:6098–6105.

28. Huang L, Forsberg CW. 1990. Cellulose digestion and cellulase regulation and distribution in *Fibrobacter succinogenes* subsp. *succinogenes* S85. Appl Environ Microbiol 56:1221–1228.
29. Béra-Maillet C, Gaudet G, Forano E. 2000. Endoglucanase activity and relative expression of glycoside hydrolase genes of *Fibrobacter succinogenes* S85 grown on different substrates. Biochim Biophys Acta 1543:77–85.
30. Holm S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6:65–70.
31. Riederer A, Takasuka TE, Makino S, Stevenson DM, Bukhman YV, Elsen NL, Fox BG. 2011. Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. Appl Environ Microbiol 77:1243–1253.
32. Stevenson DM, Weimer PJ. 2005. Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture. Appl Environ Microbiol 71:4672–4678.
33. Abdul Rahman N, Parks DH, Vanwonterghem I, Morrison M, Tyson GW, Hugenholtz P. 2015. A phylogenomic analysis of the bacterial phylum Fibrobacteres. Front Microbiol 6:1469.
34. Coen JA, Dehority BA. 1970. Degradation and utilization of hemicellulose from intact forages by pure cultures of rumen bacteria. Appl Microbiol 20:362–368.
35. Thomas S, Russell JB. 2004. The effect of cellobiose, glucose, and cellulose on the survival of *Fibrobacter succinogenes* A3C cultures grown under ammonia limitation. Curr Microbiol 48:219–223.
36. McGavin M, Forsberg CW. 1988. Isolation and characterization of endoglucanases 1 and 2 from *Bacteroides succinogenes* S85. J Bacteriol 170:2914–2922.
37. Gong J, Egbosimba EE, Forsberg CW. 1996. Cellulose-binding proteins of *Fibrobacter succinogenes* and the possible role of a 180-kDa cellulose-binding glycoprotein in adhesion to cellulose. Can J Microbiol 42:453–460.
38. Raut MP, Karunakaran E, Mukherjee J, Biggs CA, Wright PC. 2015. Influence of substrates on the surface characteristics and membrane proteome of *Fibrobacter succinogenes* S85. PLOS ONE 10:e0141197.
39. McGavin M, Lam J, Forsberg CW. 1990. Regulation and distribution of *Fibrobacter succinogenes* subsp. *succinogenes* S85 endoglucanases. Appl Environ Microbiol 56:1235–1244.
40. Burnet MC, Dohnalkova AC, Neumann AP, Lipton MS, Smith RD, Suen G, Callister SJ. 2015. Evaluating Models of Cellulose Degradation by *Fibrobacter succinogenes* S85. PLOS ONE 10:e0143809.
41. Wells JE, Russell JB, Shi Y, Weimer PJ. 1995. Cellodextrin efflux by the cellulolytic ruminal bacterium *Fibrobacter succinogenes* and its potential role in the growth of nonadherent bacteria. Appl Environ Microbiol 61:1757–1762.
42. Weimer PJ. 1993. Effects of dilution rate and pH on the ruminal cellulolytic bacterium *Fibrobacter succinogenes* S85 in cellulose-fed continuous culture. Arch Microbiol 160:288–294.
43. Shi Y, Weimer PJ. 1996. Utilization of individual cellodextrins by three predominant ruminal cellulolytic bacteria. Appl Environ Microbiol 62:1084–1088.

44. Scott HW, Dehority BA. 1965. Vitamin requirements of several cellulolytic rumen bacteria. J Bacteriol 89:1169–1175.
45. Weimer PJ, Shi Y, Odt CL. 1991. A segmented gas/liquid delivery system for continuous culture of microorganisms on insoluble substrates and its use for growth of *Ruminococcus flavefaciens* on cellulose. Appl Microbiol Biotechnol 36:178–183.
46. DuBois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. 1956. Colorimetric method for determination of sugars and related substances. Analytical Chemistry 28:350–356.
47. Miller GL, Blum R, Glennon WE, Burton AL. 1960. Measurement of carboxymethylcellulase activity. Anal Biochem 1:127–132.
48. Weimer PJ, Lopez-Guisa JM, French AD. 1990. Effect of cellulose fine structure on kinetics of its digestion by mixed ruminal microorganisms in vitro. Appl Environ Microbiol 56:2421–2429.
49. Chen J, Weimer P. 2001. Competition among three predominant ruminal cellulolytic bacteria in the absence or presence of non-cellulolytic bacteria. Microbiology (Reading, Engl) 147:21–30.
50. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.
51. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:550.
52. Ignatiadis N, Klaus B, Zaugg JB, Huber W. 2016. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nat Methods 13:577–580.
53. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological) 57:289–300.
54. Langfelder P, Zhang B, Horvath S. 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24:719–720.
55. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731.
56. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195.
57. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445-451.

Chapter 5: Conclusions

5.1 Summative discussion

Bacteria in the genus *Fibrobacter* are important primary degraders of plant cell wall polysaccharides (PCWP) in the gastrointestinal tract of diverse herbivorous animals (1). The overall objective of the research described in this thesis was to identify functional capabilities associated with certain *Fibrobacter* phylotypes that could explain the ecological patterns of abundance and distribution observed for these bacteria in their respective gastrointestinal habitats. To achieve this goal, we complemented traditional culture-based isolation and characterization with several different applications of next-generation sequence data. In chapter two, we described the development of a novel method for isolating *Fibrobacter* cultures from gastrointestinal samples, and demonstrated its utility in establishing a phylogenetically diverse collection of *Fibrobacter* strains representative of dominant populations in the herbivore gut (2). In chapter three we introduced the most extensive compilation of *Fibrobacter* genomes available to date, and performed a comparative analysis of these genomes to generate predictions of functional differences among phylotypes. Lastly, in chapter four we presented a thorough examination of global gene expression by continuous cultures of the *Fibrobacter succinogenes* type strain, S85, in response to growth rate and carbon substrate in order to gain further insights into *Fibrobacter* physiology as well as the unique mechanism of cellulose degradation employed by these bacteria.

Reports using culture-independent methods to investigate the gut microbiotas of diverse herbivores have typically identified sequences classified to the genus *Fibrobacter*, or phylum Fibrobacteres, but the relatively large amount of phylogenetic diversity estimated from the sequences recovered remains largely unexplained (3, 4). We hypothesized that host species and/or gastrointestinal tract location were important determinants of phylogenetic diversity

among *Fibrobacter* spp. in the herbivore gut and tested this hypothesis by performing a survey of diverse herbivorous hosts using both culture-based and culture-independent methods (2). The results of our survey confirmed previous reports of extensive phylogenetic diversity in the genus *Fibrobacter* (5). We identified at least 10 distinct phylotypes among our 45 *Fibrobacter* isolates and seven previously described reference strains, based on 97% nucleotide identity in the 16S rRNA gene. A subsequent whole-genome sequence-based analysis suggested that this initial estimate of diversity may be overly conservative because 21 discrete groups were identified among 40 of the *Fibrobacter* strains using a cutoff of 95% whole-genome average nucleotide identity (ANI), which has been proposed for differentiating bacterial species (6). We recovered isolates from both major lineages within the genus *Fibrobacter*, represented by the formally designated species *F. succinogenes* and *F. intestinalis* (7). Our data indicates that most of the *Fibrobacter* phylogenetic diversity present in the herbivore gut can be placed within the species *F. succinogenes*, though phylotypes of *F. intestinalis* are typically dominant in the hindguts of pigs, apes, and the ostrich. Our analysis of Fibrobacteres sequences from the Global Rumen Census (8) identified similar *Fibrobacter* phylotypes in the rumen regardless of ruminant species, evidence against the hypothesis that host species is an important contributor to *Fibrobacter* diversity, at least at the level of < 97% 16S rRNA gene identity. It remains unclear whether populations within a *Fibrobacter* phylotype, > 97% shared identity in 16S rRNA gene sequence, exhibit host association, because a robust analysis at this level of taxonomic resolution was not possible from the data analyzed here. Our survey did reveal evidence supporting a role for gastrointestinal tract location in shaping *Fibrobacter* diversity. Certain phylotypes of *F. succinogenes* were consistently observed in the rumen, while others that were common in the hindguts of horses and other large hindgut-fermenting herbivores were either

rare, or nonexistent, in the rumen. It should be emphasized that our survey was restricted to domestic and captive hosts, and, although the associations reported were strong, there likely exists additional outstanding phylogenetic diversity of *Fibrobacter* spp. in wild animal populations requiring characterization (9, 10).

A remarkably conserved physiology has been reported for *F. succinogenes* and *F. intestinalis* despite phylogenetic divergence on the order of what is typically observed for distinct bacterial genera, e.g. *Escherichia* and *Proteus* (5). We hypothesized that *Fibrobacter* spp. in the herbivore gut share a conserved role as cellulose-degrading specialists within their respective microbial communities. The data generated here is supportive of this claim because all of the *Fibrobacter* strains recovered grew readily on crystalline cellulose, but not on xylan or other monosaccharides abundant in PCWP (other than cellobiose and glucose). Although the use of cellulose for selection during isolation precludes our ability to definitively state that all *Fibrobacter* populations in the herbivore gut are cellulose degraders, our culture-independent survey suggests that our cultures are largely representative of the total phylogenetic diversity present. In addition to similarities in substrate utilization, all of our *Fibrobacter* isolates produced succinate as the major fermentation product along with lesser amounts of acetate. Succinate was also the major fermentation product produced by *F. succinogenes* S85 regardless of carbon substrate or growth rate. Potential differences were observed among *Fibrobacter* strains regarding the minor fermentation product formate. Most strains produced traces of formate (< 1 mM), but none was detected in spent media from cultures of strains in phylotypes Fs V and Fs VI (2). The failure to detect formate was associated with the absence of a gene encoding for pyruvate formate lyase (aka formate C-acetyltransferase), providing further evidence that these strains do not produce this fermentation product. It remains to be determined

whether these strains have an alternative pathway for the formation of minor fermentation products. Another notable difference observed among the *Fibrobacter* strains was variation in the ability to grow on D-lactose. Only 6 strains grew readily on this substrate, and 5 of these were of rumen origin. We failed to identify genes potentially responsible for the variation observed in this phenotype, but D-lactose is more likely to be available in the rumen than in the hindgut, particularly prior to weaning (11).

Differences in nitrogen metabolism observed among *Fibrobacter* phylotypes provided additional evidence supporting the hypothesis that gastrointestinal location impacts *Fibrobacter* evolution. Genes annotated to encode urease were identified in the genomes of 12 *Fibrobacter* strains, 10 of which were representative of phylotypes common in the hindguts of hindgut-fermenting herbivores, *e.g.* horses, elephants, and rhinos. Growth assays confirmed that possession of these genes was associated with an ability to utilize urea as a sole source of nitrogen for growth. Important differences exist between the rumen and the hindgut regarding the form of nitrogen-containing compounds available to the microbiota. In the rumen, microbes have access to dietary sources of nitrogen prior to these compounds being absorbed by the host. In contrast, substrates available to the hindgut microbiota have already passed through the small intestine where dietary protein is degraded and absorbed. As a result, microbes living in the lower gut are relatively more dependent on endogenous sources of nitrogen such as urea, which is synthesized by the host and transferred through the circulatory system to the hindgut (12).

The ecological patterns observed for *Fibrobacter* phylotypes is strongly suggestive of their conserved niche as primary degraders of PCWP in the herbivore gut. Our culture-independent analysis using 16S rRNA gene amplicon sequencing identified relatively abundant (1 - 0.1%) *Fibrobacter* populations in the rumen as well as the hindguts of large herbivores.

Less abundant *Fibrobacter* populations were observed in the hindguts of omnivores and herbivores that generally consume a diet that is more readily digestible, *e.g.* pigs and primates. *Fibrobacter* populations were not observed in samples collected from carnivores. Taken together, these results suggest an association between the overall fiber-degrading capacity of a gut microbiota (13) and the relative abundance of *Fibrobacter* within that community. Moreover, evidence was obtained suggesting a link between the fibrolytic capacity of the gut microbiota and the specific *Fibrobacter* phylotypes present. Phylotypes of *F. succinogenes* tended to predominate in the rumen as well as in the hindguts of other large, strict herbivores, namely horses, rhinos, tapirs and elephants. Less fibrolytic gut microbiotas typically had relatively more abundant populations of *F. intestinalis* phylotypes. Whole-genome sequencing identified substantial differences in the numbers of genes annotated to carbohydrate-active enzymes (CAZymes) (14) among the *Fibrobacter* species that may provide a functional explanation for the ecological distributions observed.

Strains of *F. succinogenes* had considerably higher numbers of total CAZyme genes as well as genes annotated as carbohydrate esterases (CEs), glycoside hydrolases (GHs), polysaccharide lyases (PLs), and carbohydrate-binding modules compared to strains of *F. intestinalis*. Enzymes from these CAZyme classes target specific carbohydrate linkages present in the plant cell wall, and their complementary activities are required for complete depolymerization. The differences in CAZyme gene content between *F. succinogenes* and *F. intestinalis* were most dramatic in enzymes families that are involved in degrading hemicellulose. A general increase in CAZyme gene content was also observed among phylotypes of *F. succinogenes* that are dominant in the rumen over phylotypes of *F. succinogenes* that were more commonly observed in the horse hindgut. CAZyme families that

were particularly enriched in rumen phylotypes of *F. succinogenes* included CBM35, CE6, GH10, GH43, and GH9. All of these enzyme families, except GH9, are utilized by *Fibrobacter* to target hemicellulose for degradation (15). It is not that surprising that CAZyme gene content is so closely associated the ecology and phylogeny of these fiber-degrading bacteria as differences in substrate preferences have been hypothesized to contribute to evolutionary diversification in other glycan utilizers present in the gut, such as species of *Bacteroides* and *Bifidobacterium* (16–18). However, in those bacteria differentiation appears to be driven more by the presence and absence of specific enzyme families, while the largest differences observed here for *Fibrobacter* involved CAZyme families that were present across all strains but varied considerably in copy number. Therefore, I hypothesize that these difference are reflective of optimization for a specific rate of fiber-degradation in their native environment as opposed to differences in substrate preferences. A larger sample of more closely related strains and species is necessary to determine whether substrate preference makes a larger contribution to intra-phylotype diversification than was observed between the phylotypes examined here.

Fibrobacter utilize a unique and poorly understood mechanism for depolymerizing PCWP (15, 19). Our results support previous observations that this strategy is generally conserved among diverse *Fibrobacter* strains (20). All of the CAZyme families previously described as being important for PCWP degradation by *Fibrobacter* were present in the genomes of the *Fibrobacter* strains analyzed here (15), including the major GH families *Fibrobacter* uses to degrade cellulose (GH5, GH8, GH9, and GH45) and hemicellulose (GH10, GH26, and GH43). No known homologs of genes encoding cellulosomal components or processive exocellulases, features characteristic of other cellulose-degrading microbes (21, 22), were observed. Another feature of the *Fibrobacter* fiber-degrading system that was conserved is the

general absence of obvious carbohydrate-binding modules associated with cellulases, while hemicellulases typically contain associated CBM6 and CBM35 domains. A notable exception is the major cellulose-binding endoglucanase described from *F. succinogenes* S85, Cel51A (23–25). This cellulase consists of a GH51 catalytic domain with an accompanying CBM11 domain. Several observations from this work support a fundamental role for this protein in *Fibrobacter* biology. A single predicted ortholog of the gene encoding Cel51A was observed in all of the *Fibrobacter* genomes analyzed here. This is atypical for CAZyme families encoding endoglucanases in *Fibrobacter*, because multiple copies of these GH families are commonly present in the genome. Moreover, our analysis of global gene expression by *F. succinogenes* S85 identified the gene encoding Cel51A as the most highly expressed CAZyme gene in the S85 genome as well as provided strong evidence that its expression is constitutive. These observations suggest that the expression of this cellulase is finely tuned because we found no evidence of gene duplication, and that there are few environments where these bacteria exist where the production of this protein is not advantageous as we found no evidence of repression.

The tight control over the expression of genes encoding hemicellulases by *F. succinogenes* S85 lends further support to the hypothesis that these CAZyme families play a complementary role in fiber degradation by *Fibrobacter* (15). Genes encoding enzymes in CAZyme families CBM6, CBM35, CE6, GH10, and GH43 are believed to be primarily involved in hemicellulose degradation by *Fibrobacter* as a means to gain access to cellulose in the plant cell wall. Elevated expression of these genes was observed only during slow growth of S85 on soluble sugars, strongly suggesting that these CAZyme families and the enzymes they encode are not required for growth on cellulose by this bacterium. It is notable that genes in these CAZyme families were also among those that displayed the most variation in gene copy number across

genomes. One proposed mechanism of niche differentiation that influences subsequent evolutionary diversification in bacteria is a change in gene expression (26). Figure 5.1 shows a plot of the variance in gene expression observed for these families for *F. succinogenes* S85 against variation in genome content of conserved CAZyme families (CBMs, CEs, GHs, and PLs) involved in PCWP degradation among diverse *Fibrobacter* strains. A moderate, but statistically significant ($P < 0.001$), positive association (rho = 0.58) was observed suggesting that degradative enzymes exhibiting greater regulation in their expression may be particularly influential in the evolution of *Fibrobacter*. This hypothesis should be explored further in future studies.

Microbial communities are incredibly diverse, and a fundamental understanding of this diversity is necessary in order to reliably manipulate these communities towards a specific purpose or application. Through the research described in this thesis, I have attempted to construct a rationale explaining the phylogenetic diversity of the fiber-degrading genus *Fibrobacter* present in the herbivore gut by examining patterns of distribution and abundance as well as the functional potential of representatives of common *Fibrobacter* phylotypes. These bacteria make an important contribution to overall gut function in their herbivorous hosts as primary degraders of PCWP, and their unique cellulolytic capability and succinate production makes them of general interest for industrial purposes. It is my hope that the observations described here, as well as the *Fibrobacter* strains and genomes generated, prove useful to future efforts aimed at furthering our knowledge of the biology of these mysterious microbes.

5.2 Future prospects

The work presented in this thesis raises several outstanding research questions worthy of future investigation. Does CAZyme gene relative abundance in a *Fibrobacter* genome correlate

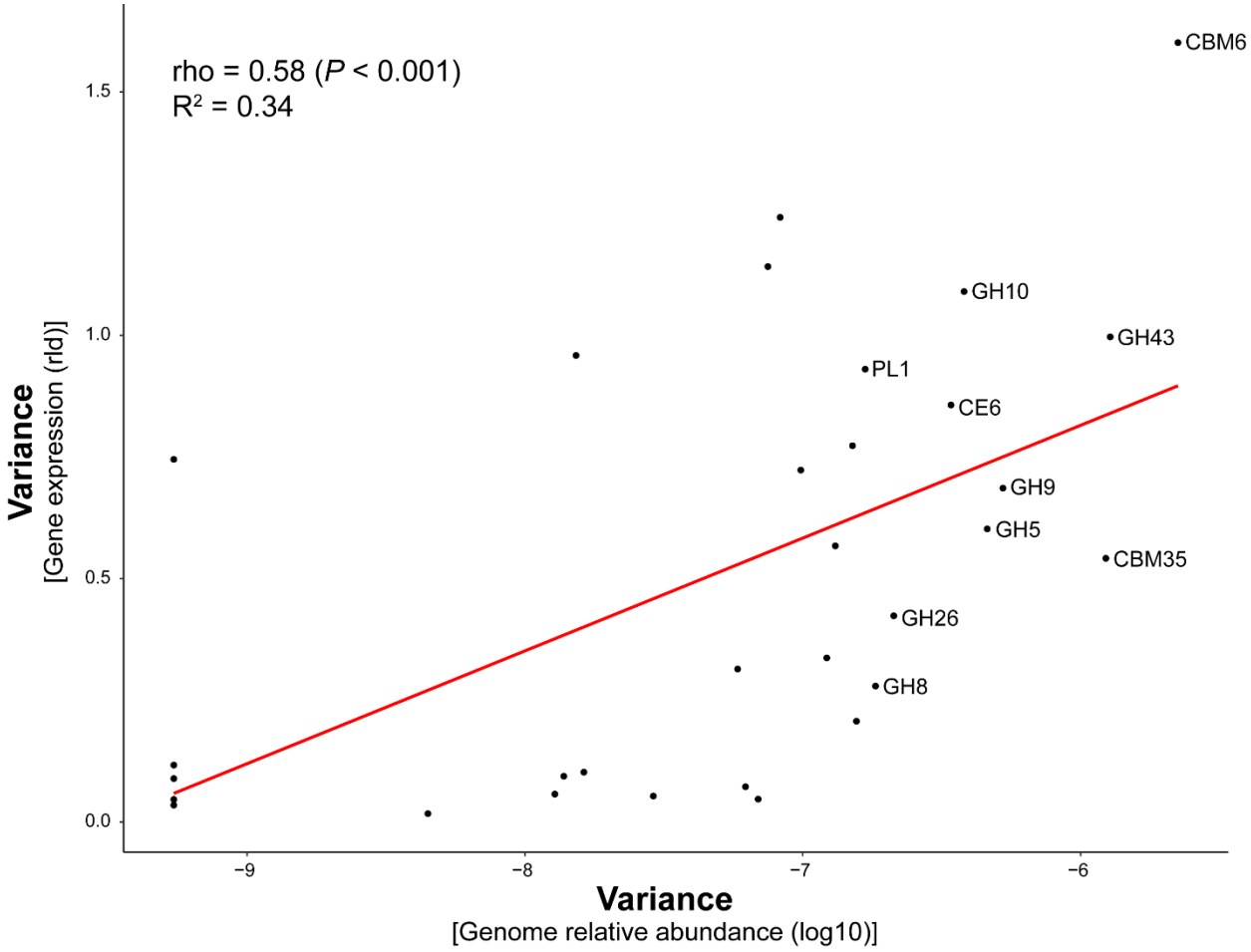


Figure 5.1: Correlation between variation in genome relative abundance and variation in gene expression for CAZyme families involved in PCWP degradation (CBMs, CEs, GHs, and PLs) by *Fibrobacter*. Variation in genome relative abundance (x-axis) is represented by the log10 transformed variance in relative abundance for a CAZyme family across the 40 *Fibrobacter* genomes (Chapter 3). Variation in gene expression (y -axis) is represented by the variance in mean regularized log2 expression (rld) for genes in a CAZyme family observed across different substrates and growth rates for *F. succinogenes* S85 (Chapter 4). Each point represents an indidual CAZyme family, with the 10 families exibiting the highest amount of variance in relative abundance across *Fibrobacter* strains labeled.

with the rate of PCWP solubilization? Do *Fibrobacter* strains target specific parts of the plant cell wall, and, if so, is this evident in their phylogeny or ecology? Is the expression of genes encoding hemicellulases tightly regulated in all *Fibrobacter* strains, and if so, are the expression

patterns observed in *F. succinogenes* S85 conserved? Do *Fibrobacter* spp. form specific associations with certain microbial taxa in the gut microbiota? What is the significance of the relatively large degree of phylogenetic diversity, based on 95% ANI, identified among these bacteria, and is this truly indicative of ecologically discrete populations? For the remainder of this section I will introduce specific hypotheses related to these questions and suggestions for testing them.

Does CAZyme gene relative abundance in the genome correlate with the rate of PCWP solubilization?

Based on the strong association observed between CAZyme genome content and the fiber-degrading potential of the gut microbiota where certain *Fibrobacter* phylotypes are found, I have proposed that higher genomic proportions of CAZyme gene families involved in PCWP degradation support higher rates of fiber-degradation. To test the validity of this hypothesis, *Fibrobacter* strains could be cultured on complex substrates/forages (*e.g.* alfalfa hay, timothy grass, orchard grass, etc.) and cellulose in order to quantify the rates at which insoluble fiber is degraded, soluble sugars are produced, and the maximum growth rate achieved by the various *Fibrobacter* strains. This would allow for the identification of differences among phylotypes as well as provide stronger evidence for a causal relationship with CAZyme genome content and fiber-degrading potential. These data as well as the maximum rate of crystalline cellulose degradation achievable for each strain would also be valuable for statistical modeling to identify novel genomic elements associated with higher rates of degradation.

Do *Fibrobacter* strains target specific parts of the plant cell wall, and, if so, is this evident in their phylogeny or ecology?

The common coexistence of multiple discrete *Fibrobacter* phylotypes within a given gut microbiota leads me to hypothesize that these phylotypes occupy distinct ecological niches *in vivo* and that this, at least partly, involves differential colonization on the surface of plant fibers. This claim could be tested by setting up co-cultures on complex substrates to determine whether strains from the different phylotypes tend to compete or coexist on various substrates. The distinct cell morphology of *Fibrobacter* phylotype Fs I (5) would allow for these strains to be readily distinguished from strains of other *Fibrobacter* phylotypes via microscopy. Probes targeting nucleic acid sequences of specific phylotypes could be used to monitor the different populations using either qPCR, fluorescent in situ hybridization, or both. Differences in substrate preferences among phylotypes could also be investigated using LC-MS (27) to determine which soluble sugars are produced, and at what rate, by strains of the different phylotypes during growth on complex substrates.

Is the expression of genes encoding hemicellulases tightly regulated in all *Fibrobacter* strains, and, if so, are the expression patterns observed in *F. succinogenes* S85 conserved?

Fibrobacter strains did not utilize monosaccharides typically found in the hemicellulosic fraction of PCWP, supporting a conserved complementary role for hemicellulases in fiber-degradation by *Fibrobacter*. Therefore, I would expect that the genes encoding CAZyme families involved in hemicellose degradation by *Fibrobacter*, which are broadly shared among strains, to be tightly regulated in a manner similar to what was observed for *F. succinogenes*

strain S85. However, additional empirical evidence supporting this claim is necessary to confirm this. An important caveat to the expression analysis performed here is that RNA content is often only moderately correlated with protein content in bacteria (28). Therefore, I would focus future efforts examining the regulation of enzyme expression in *Fibrobacter* at the protein level. It would also be beneficial to examine specific enzymatic activities of cells and supernatants from cultures grown on different substrates and at different growth rates. For example, testing whether diverse *Fibrobacter* cultures growing slowly on soluble sugars consistently produce more xylanase activity than when they are cultured on cellulose.

Do *Fibrobacter* phylotypes form specific associations with certain microbial taxa in the gut microbiota?

One important aspect of *Fibrobacter* ecology that was not explored in depth in this thesis is interactions with other microbial species of the gut microbiota. Stable co-cultures of *Fibrobacter* with diverse non-cellulolytic bacteria have been reported, and it is believed that, at least under these conditions, both organisms benefit from the association (29–32). I hypothesize that *Fibrobacter* spp. do engage in mutualistic relationships with other microorganisms in the gut and that these relationships have important implications for *Fibrobacter* ecology and evolution. This could be tested by mining data from culture-independent surveys of herbivore gut microbial communities and building an interaction network to determine cooccurrence. A large dataset would likely be required to identify robust associations due to the complexity of these gut microbial communities.

Moreover, anaerobic cellulose-degrading bacteria often form close associations with methanogens that consume the H_2 produced by these bacteria (33, 34). *F. succinogenes* S85

does not produce H_2 and it is assumed that this feature makes cellulose-utilization by *Fibrobacter* in the rumen preferential because less potential energy is lost to methane production (35). However, *F. succinogenes* S85 does produce small amounts of formate and this product can support methanogenic growth in coculture (36). We identified *Fibrobacter* phylotypes that likely do not produce formate because none was detected in spent culture medium, nor do these strains possess the gene encoding pyruvate formate lyase. It would be interesting to determine whether *Fibrobacter* strains from these phylotypes are capable of supporting methanogens in coculture, and if so, to determine the product they are producing that is being utilized by the methanogens. Although H_2 production has not been reliably demonstrated for any *Fibrobacter* strains, we did not test our strains for this capability.

What is the significance of the large amount of phylogenetic diversity identified, based on 95% ANI, and is this diversity representative ecologically discrete populations in the gut?

We identified a surprising amount of phylogenetic diversity among our *Fibrobacter* genomes, particularly considering the few phenotypic differences observed among the strains. However, due to the substantial complexity of carbohydrates present in PCWP, I believe that there is ample opportunity for niche differentiation among these bacteria and that they do occupy subtly different ecological niches *in vivo* (37). *F. succinogenes* phylotypes Fs I and Fs II, particularly those living in the rumen of cattle, represent good targets to investigate this hypothesis further. Strains from these phylotypes are relatively easy to isolate from rumen samples using the techniques described in this thesis. Consequently, a large collection of very closely related strains representative of these phylotypes should be readily obtainable. This would allow for a much more robust analysis of potential genomic differences suggestive of

distinct ecologies. An alternative to a culture-based approach would be to use metagenomics, or qPCR, to perform higher resolution surveys of sequence-discrete populations in the rumen over time, and in response to different environment impacts (*e.g.* diet and age) in order to identify stable patterns of abundance and distribution capable of differentiating distinct species-level groups within phylotypes.

5.3 References

1. Stewart CS, Flint HJ. 1989. *Bacteroides* (*Fibrobacter*) *succinogenes*, a cellulolytic anaerobic bacterium from the gastrointestinal tract. Appl Microbiol Biotechnol 30:433–439.
2. Neumann AP, McCormick CA, Suen G. 2017. *Fibrobacter* communities in the gastrointestinal tracts of diverse hindgut-fermenting herbivores are distinct from those of the rumen. Environ Microbiol 19:3768–3783.
3. Jewell KA, Scott JJ, Adams SM, Suen G. 2013. A phylogenetic analysis of the phylum Fibrobacteres. Syst Appl Microbiol 36:376–382.
4. Ransom-Jones E, Jones DL, Edwards A, McDonald JE. 2014. Distribution and diversity of members of the bacterial phylum Fibrobacteres in environments where cellulose degradation occurs. Syst Appl Microbiol 37:502–509.
5. Amann RI, Lin C, Key R, Montgomery L, Stahl DA. 1992. Diversity among *Fibrobacter* isolates: towards a phylogenetic classification. Systematic and Applied Microbiology 15:23–31.
6. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. 2015. Microbial species delineation using whole genome sequences. Nucleic Acids Res 43:6761–6771.
7. Montgomery L, Flesher B, Stahl D. 1988. Transfer of *Bacteroides succinogenes* (Hungate) to *Fibrobacter* gen. nov. as *Fibrobacter succinogenes* comb. nov. and description of *Fibrobacter intestinalis* sp. nov. International Journal of Systematic and Evolutionary Microbiology 38:430–435.
8. Henderson G, Cox F, Ganesh S, Jonker A, Young W, Collaborators GRC, Abecia L, Angarita E, Aravena P, Arenas GN, Ariza C, Attwood GT, Avila JM, Avila-Stagno J, Bannink A, Barahona R, Batistotti M, Bertelsen MF, Brown-Kav A, Carvajal AM, Cersosimo L, Chaves AV, Church J, Clipson N, Cobos-Peralta MA, Cookson AL, Cravero S, Carballo OC, Crosley K, Cruz G, Cucchi MC, Barra R de la, Menezes ABD, Detmann E, Dieho K, Dijkstra J, Reis WLS dos, Dugan MER, Ebrahimi SH, Eythórsdóttir E, Fon FN, Fraga M, Franco F, Friedeman C, Fukuma N, Gagić D, Gangnat I, Grilli DJ, Guan LL, Miri VH, Hernandez-Sanabria E, Gomez AXI, Isah OA, Ishaq S, Jami E, Jelincic J, Kantanen J, Kelly WJ, Kim S-H, Klieve A, Kobayashi Y, Koike S, Kopecny J, Kristensen TN, Krizsan SJ, LaChance H, Lachman M, Lamberson WR, Lambie S, Lassen J, Leahy SC, Lee S-S, Leiber F, Lewis E, Lin B, Lira R, Lund P, Macipe E, Mamuad LL, Mantovani HC, Marcoppido GA, Márquez C, Martin C, Martinez G, Martinez ME, Mayorga OL, McAllister TA, McSweeney C, Mestre L, Minnee E, Mitsumori M, Mizrahi I, Molina I, Muenger A, Muñoz C, Murovec B, Newbold J, Nsereko V, O'Donovan M, Okunade S, O'Neill B, Ospina S, Ouwerkerk D, Parra D, Pereira LGR, Pinares-Patiño C, Pope PB, Poulsen M, Rodehutscord M, Rodriguez T, Saito K, Sales F, Sauer C, Shingfield K, Shoji N, Simunek J, Stojanović-Radić Z, Stres B, Sun X, Swartz J, Tan ZL, Tapio I, Taxis TM, Tomkins N, Ungerfeld E, Valizadeh R, Adrichem P van, Hamme JV, Hoven WV, Waghorn G, Wallace RJ, Wang M, Waters SM, Keogh K, Witzig M, Wright A-DG, Yamano H, Yan T, Yáñez-Ruiz DR, Yeoman CJ, Zambrano R, Zeitz J, Zhou M, Zhou HW, Zou CX, Zunino P, 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.

9. Metcalf JL, Song SJ, Morton JT, Weiss S, Seguin-Orlando A, Joly F, Feh C, Taberlet P, Coissac E, Amir A, Willerslev E, Knight R, McKenzie V, Orlando L. 2017. Evaluating the impact of domestication and captivity on the horse gut microbiome. Scientific Reports 7:15497.
10. McKenzie VJ, Song SJ, Delsuc F, Prest TL, Oliverio AM, Korpita TM, Alexiev A, Amato KR, Metcalf JL, Kowalewski M, Avenant NL, Link A, Di Fiore A, Seguin-Orlando A, Feh C, Orlando L, Mendelson JR, Sanders J, Knight R. 2017. The effects of captivity on the mammalian gut microbiome. Integr Comp Biol 57:690–704.
11. Ghali I, Sofyan A, Ohmori H, Shinkai T, Mitsumori M. 2017. Diauxic growth of *Fibrobacter succinogenes* S85 on cellobiose and lactose. FEMS Microbiol Lett 364.
12. Bergen WG, Wu G. 2009. Intestinal nitrogen recycling and utilization in health and disease. J Nutr 139:821–825.
13. Van Soest PJ. 1996. Allometry and ecology of feeding behavior and digestive capacity in herbivores: A review. Zoo Biology 15:455–479.
14. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37:D233-238.
15. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D, Brumm PJ. 2011. The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. PLoS ONE 6:e18814.
16. Milani C, Turroni F, Duranti S, Lugli GA, Mancabelli L, Ferrario C, van Sinderen D, Ventura M. 2016. Genomics of the genus *Bifidobacterium* reveals species-specific adaptation to the glycan-rich gut environment. Appl Environ Microbiol 82:980–991.
17. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL. 2010. Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations. Cell 141:1241–1252.
18. Xu J, Mahowald MA, Ley RE, Lozupone CA, Hamady M, Martens EC, Henrissat B, Coutinho PM, Minx P, Latreille P, Cordum H, Van Brunt A, Kim K, Fulton RS, Fulton LA, Clifton SW, Wilson RK, Knight RD, Gordon JI. 2007. Evolution of symbiotic bacteria in the distal human intestine. PLoS Biol 5:e156.
19. Wilson DB. 2009. Evidence for a novel mechanism of microbial cellulose degradation. Cellulose 16:723–727.
20. Abdul Rahman N, Parks DH, Vanwonterghem I, Morrison M, Tyson GW, Hugenholtz P. 2015. A phylogenomic analysis of the bacterial phylum Fibrobacteres. Front Microbiol 6:1469.
21. Teeri TT. 1997. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Trends in Biotechnology 15:160–167.
22. Artzi L, Bayer EA, Moraïs S. 2017. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol 15:83–95.
23. McGavin M, Forsberg CW. 1988. Isolation and characterization of endoglucanases 1 and 2 from *Bacteroides succinogenes* S85. J Bacteriol 170:2914–2922.
24. McGavin M, Forsberg CW. 1989. Catalytic and substrate-binding domains of endoglucanase 2 from *Bacteroides succinogenes*. J Bacteriol 171:3310–3315.

25. Gong J, Egbosimba EE, Forsberg CW. 1996. Cellulose-binding proteins of *Fibrobacter succinogenes* and the possible role of a 180-kDa cellulose-binding glycoprotein in adhesion to cellulose. Can J Microbiol 42:453–460.
26. Deng J, Auchtung J, Konstantinidis KT, Caro-Quintero A, Brettar I, Höfle M, Tiedje JM. 2017. Divergence in gene regulation contributes to sympatric speciation of *Shewanella balticastrains*. Appl Environ Microbiol.
27. Arntzen MØ, Várnai A, Mackie RI, Eijsink VGH, Pope PB. 2017. Outer membrane vesicles from *Fibrobacter succinogenes* S85 contain an array of carbohydrate-active enzymes with versatile polysaccharide-degrading capacity. Environ Microbiol 19:2701–2714.
28. Vogel C, Marcotte EM. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13:227–232.
29. Sawanon S, Koike S, Kobayashi Y. 2011. Evidence for the possible involvement of *Selenomonas ruminantium* in rumen fiber digestion. FEMS Microbiol Lett 325:170–179.
30. Chen J, Weimer P. 2001. Competition among three predominant ruminal cellulolytic bacteria in the absence or presence of non-cellulolytic bacteria. Microbiology (Reading, Engl) 147:21–30.
31. Scheifinger CC, Wolin MJ. 1973. Propionate formation from cellulose and soluble sugars by combined cultures of *Bacteroides succinogenes* and *Selenomonas ruminantium*. Appl Microbiol 26:789–795.
32. Stanton TB, Canale-Parola E. 1980. *Treponema bryantii* sp. nov., a rumen spirochete that interacts with cellulolytic bacteria. Arch Microbiol 127:145–156.
33. Weimer PJ, Zeikus JG. 1977. Fermentation of cellulose and cellobiose by *Clostridium thermocellum* in the absence of *Methanobacterium thermoautotrophicum*. Appl Environ Microbiol 33:289–297.
34. Pavlostathis SG, Miller TL, Wolin MJ. 1990. Cellulose fermentation by continuous cultures of *Ruminococcus albus* and *Methanobrevibacter smithii*. Appl Microbiol Biotechnol 33:109–116.
35. Chaucheyras-Durand F, Masséglia S, Fonty G, Forano E. 2010. Influence of the composition of the cellulolytic flora on the development of hydrogenotrophic microorganisms, hydrogen utilization, and methane production in the rumens of gnotobiotically reared lambs. Appl Environ Microbiol 76:7931–7937.
36. Rychlik JL, May T. 2000. The effect of a methanogen, *Methanobrevibacter smithii*, on the growth rate, organic acid production, and specific ATP activity of three predominant ruminal cellulolytic bacteria. Curr Microbiol 40:176–180.
37. Hamaker BR, Tuncil YE. 2014. A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. J Mol Biol 426:3838–3850.

Appendices

Appendix 1: Influence of growth-promoting antibiotics on bacterial populations in the intestines of broiler chickens

Authors and contributors:

Anthony Neumann: Designed the experiments, performed the experiments, analyzed the data, and wrote the manuscript

Dr. Garret Suen: Designed the experiments, and edited the manuscript

Adapted from:

Neumann, A. P., and Suen, G. Differences in major bacterial populations in the intestines of mature broilers after feeding virginiamycin or bacitracin methylene disalicylate. J Appl Microbiol. 2015 Dec;119(6):1515-26.

Funding:

This work was financially supported through industrial grants by Pfizer Animal Health Global Poultry and Zoetis Animal Health, LLC. Zoetis is the licensed owner and distributor of BMD. We confirm that we have no direct connection, or personal interest, in the products being evaluated in this study.

A1.1 Abstract

Commercial poultry producers in the United States commonly supplement broiler diets with antibiotics in order to improve intestinal health and production performance, however it is unclear which specific effects on the GI microbiota lead to improvements in performance. In this study, we examined the influence of two commonly used antimicrobials, virginiamycin and bacitracin methylene disalicylate (BMD), on bacterial communities in the GI tract of market-age broiler chickens over four replicate grow-outs using 454 pyrosequencing and *Clostridium perfringens* specific quantitative PCR. Our results suggest that virginiamycin has a more pronounced impact on broiler GI tract bacterial communities, relative to BMD, and that this effect manifests itself primarily through significant enrichments in the genus *Faecalibacterium* in the cecum as well as a distinct population of *Lactobacillus*, OTU_02, in both the jejunum and cecum. No significant alterations in major bacterial populations were observed after feeding BMD, but its inclusion in broiler diets did produce a numerical increase in another abundant *Lactobacillus* population, OTU_01, particularly versus the virginiamycin treatment. Contrary to what was originally predicted, no evidence for a difference among the diets in the levels of *C. perfringens* in the jejunum or cecum was observed. However, our data did identify a potential link between the interval between grow-outs and *C. perfringens* abundance in the GI tract.

A1.2 Introduction

Supplementation of livestock diets with United States Food and Drug Administration-approved antibiotics is widely used as a means to promote gastrointestinal (GI) health and production performance (1, 2). These effects are thought to involve GI microbes because germ-free chicks do not show growth rate improvements when antibiotics are administered through their feed (3). Two antibiotics commonly included in commercial broiler feed are virginiamycin and bacitracin methylene disalicylate (BMD). Both of these compounds preferentially target Gram-positive bacteria, but their antibacterial effects occur via entirely different mechanisms (4). Virginiamycin is a streptogramin consisting of two components that act synergistically to irreversibly inhibit bacterial protein synthesis by binding to the 23S rRNA of the 50S ribosomal subunit (4). Bacitracin is a mixture of nonribosomal polypeptides that interfere with formation of the bacterial cell wall by preventing dephosphorylation of the carrier for N-acetylmuramyl pentapeptide intermediates (4).

Several studies have reported alterations in the GI tract bacterial community of broiler chickens associated with dietary supplementation with either virginiamycin or bacitracin (5–8), but rarely have these effects been assessed over consecutive flocks. Recently, a feeding trial was conducted to examine the effect of feeding virginiamycin or BMD over multiple flocks, or grow-out cycles, of broilers raised under typical commercial conditions (9). Although significantly better feed efficiency (*i.e.*, feed consumed per weight gained) was associated with virginiamycin, compared to BMD during the first grow-out, a trend toward better performance associated with BMD supplementation was evident by the third grow-out. The authors postulated that these effects could be explained, in part, by outgrowth of opportunistic pathogens associated with significantly higher mortality in the virginiamycin-fed broilers during the third grow-out because

mortality coincided with significantly higher levels of *Clostridium perfringens* in the litter of those pens.

C. perfringens are spore-forming, obligately anaerobic bacteria recognized as the causative agent of necrotic enteritis (NE) in avians (10, 11). NE is a significant concern for poultry producers world-wide with annual economic losses estimated to be in excess of two billion U.S. dollars (12). Despite the widespread prevalence of NE, in-feed antibiotics have been associated with protection of flocks from the disease (13). Both virginiamycin and bacitracin have been reported to reduce the incidence of NE as well as *C. perfringens* levels in the GI tract (14–16). However, only one study has specifically examined the effects of feeding virginiamycin or bacitracin to broilers over consecutive grow-outs and provided evidence for a difference in how these compounds impact *C. perfringens* levels over time (9).

In this study, jejunal and cecal contents from market-age broiler chickens raised on diets supplemented with either virginiamycin or bacitracin were examined using 454 pyrosequencing and *C. perfringens* specific quantitative PCR (qPCR) in order to investigate differences in GI bacterial community composition across four replicate grow-outs performed consecutively. The jejunum and cecum were chosen as the GI sites to be investigated because of their distinct roles in host digestive physiology and associated bacterial communities (17–20). Observations from previous work (9) led us to hypothesize that virginiamycin would have a broader impact, compared with BMD, on GI bacterial populations important for protecting the host from the outgrowth of opportunistic pathogens, *e.g. C. perfringens*. Our results present evidence that virginiamycin has a greater influence on GI bacterial populations than BMD, and suggest that there are differences in how these two dietary supplements influence the abundance of specific bacterial populations in the broiler GI tract. Moreover, the sequencing depth generated from the

120 individual birds sampled makes this work the highest resolution investigation of broiler GI bacterial populations in response to these antimicrobials available to date.

A1.3 Results

Overall, 120 Cobb 500 broilers were sampled to assess the impact of feeding virginiamycin or BMD on the resident bacterial community composition within the GI tract over four replicate grow-outs. To investigate these effects, we performed a combination of 454 pyrosequencing and *C. perfringens* targeted qPCR on jejunal and cecal contents from these 120 broilers.

Total Bacteria present in intestinal contents of mature broiler chickens

A grand total of 1,702,276 V6-V8 16S rRNA gene sequences, 983,627 from jejunal contents and 718,649 from cecal contents, were recovered after quality control and processing. The minimum number of processed sequences in a single dietary treatment replicate was 50,408. Sample coverages and community richness estimates, based on the total number of operational taxonomic units (OTUs) observed in a single treatment replicate, can be found in Table A1.1. Average Good's coverage for jejunal bacterial communities from a single treatment replicate ranged from 0.99 to approximately 1.00. The mean number of OTUs in the jejunum among these groups ranged from 434.45 (virginiamycin rep 3) to 901.07 (control rep 4). Average Good's coverage for cecal bacterial communities ranged from 0.94 to 0.96. The mean number of OTUs per group in the cecum ranged from 3,793.89 (virginiamycin rep 3) to 5,170.30 (control rep 4). 454 pyrosequencing also allowed for the identification and approximate quantification of the most prevalent bacterial groups in jejunal and cecal contents from 6 week old broiler

Table A1.1: Sequence metrics and sample diversity

Diet	Rep	Jejunum			Cecum		
		OTUs per Sample	Good's Coverage	Shannon diversity (*H'*)	OTUs per Sample	Good's Coverage	Shannon diversity (*H'*)
Control	1	606.95	0.99	2.00	5093.14	0.94	5.83
Control	2	553.20	0.99	2.27	3903.34	0.96	5.58
Control	3	680.63	0.99	2.74	4015.45	0.96	5.77
Control	4	901.07	0.99	2.35	5170.30	0.94	5.78
Mean	-	685.46	0.99	2.34	4545.56	0.95	5.74[a]
Virginiamycin	1	687.33	0.99	2.31	4652.16	0.95	5.35
Virginiamycin	2	527.96	0.99	2.26	4151.57	0.96	5.62
Virginiamycin	3	434.45	1.00	2.05	3793.89	0.96	5.32
Virginiamycin	4	771.75	0.99	2.02	4337.00	0.95	5.49
Mean	-	605.37	0.99	2.16	4233.66	0.95	5.44[b]
BMD	1	854.41	0.99	2.46	4920.25	0.94	5.57
BMD	2	749.91	0.99	2.42	4377.49	0.95	5.62
BMD	3	538.78	0.99	2.39	3930.49	0.96	5.57
BMD	4	787.38	0.99	2.07	4819.57	0.94	5.82
Mean	-	732.62	0.99	2.33	4511.95	0.95	5.64[ab]

Values represent averages obtained from 1,000 simulations for each group after normalizing to 50,408 sequences per group

[a-b]Different letters indicate significant differences among the means in the same column ($P < 0.05$)

chickens and confirmed that these two anatomical sites harbor distinct bacterial communities (Figure A1.1). ANOSIM, based on Bray-Curtis dissimilarity, provided statistical evidence for a highly significant difference between the bacterial community structure in jejunal versus cecal contents as well ($P < 0.001$).

Dominant taxa present in intestinal contents of mature broiler chickens

The most abundant phyla observed in jejunal contents were the Firmicutes followed by Actinobacteria. The relative abundances of Firmicutes in jejunal contents among the treatment

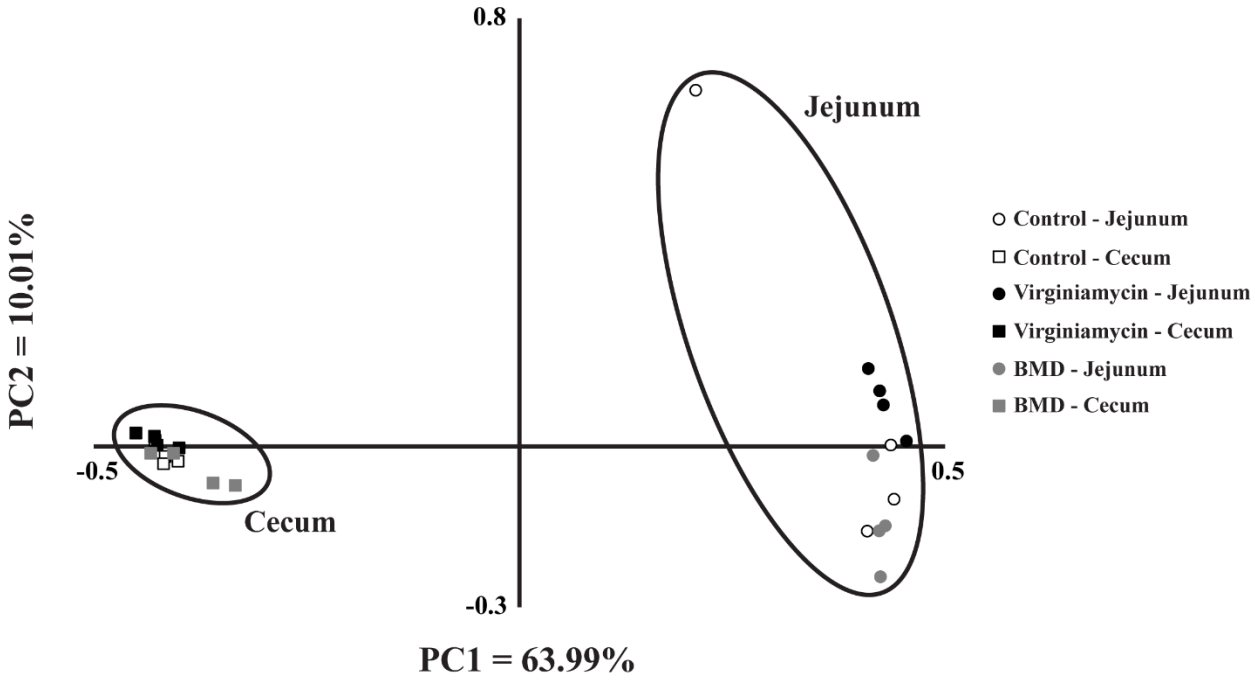


Figure A1.1: Relationships among total bacterial communities from each dietary treatment replicate. The first two dimensions of principal component analysis (PCoA) of Bray-Curtis dissimilarity among bacterial communities from jejunal contents, plotted with circles (○), and cecal contents, plotted with squares (□), are represented by the scatterplot. Circles are also drawn around the communities originating from the same GI section, and labeled accordingly. The percent of total variation explained by each axis can be found in the axis labels. Open shapes indicate the control diet communities, black filled shapes the virginiamycin communities, and grey filled shapes the BMD communities.

replicates ranged from 95.07% (control rep 3) to 99.77% (virginiamycin rep 3). Actinobacteria were considerably less abundant with relative abundances between 0.13% (virginiamycin rep 3) and 4.72% (control rep 3). Of the ten major OTUs that made up on average at least one percent of the bacterial community in the jejunum, six were classified to the genus *Lactobacillus* with high confidence and one additional OTU was classified as belonging to the family Lactobacillales (Figure A1.2A). Two distinct groups of *Lactobacillus*, OTU_01 and OTU_02, accounted for on average 48.60% of all bacteria present in the jejunum across all treatments and replicates. BLAST analysis of representative sequences for these OTUs revealed closest GenBank matches of *Lactobacillus crispatus* strain BC5 and *Lactobacillus johnsonii* strain

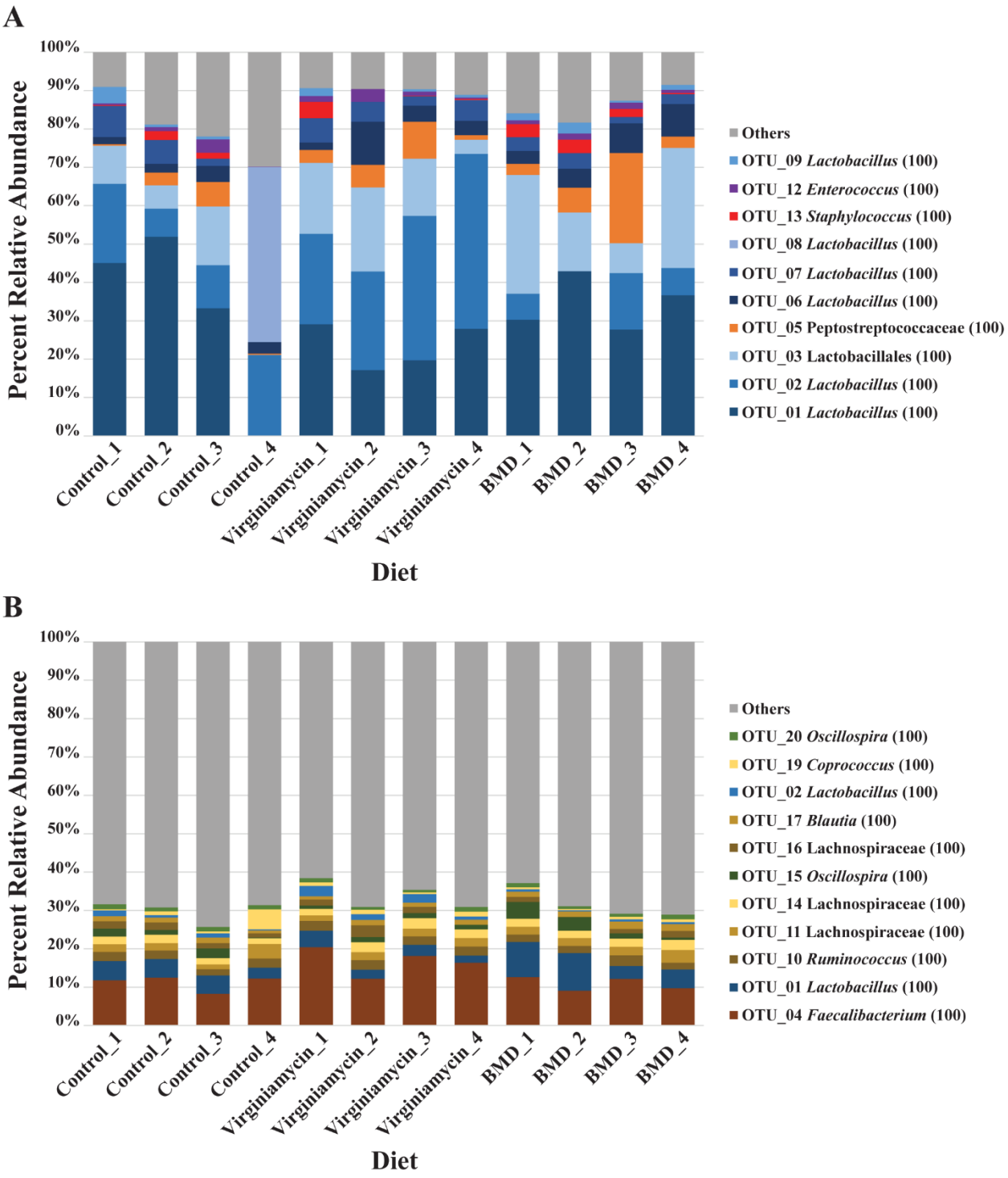


Figure A1.2: Major bacterial OTUs among the dietary treatments and replicates (average percent relative abundance of ≥ 1%). OTUs observed at an average percent relative abundance of ≤ 1% represented in the category "Others". The closest taxonomic classification for each OTU is indicated, with consensus confidence/bootstrap value in parentheses. Subsampling to 50,408 sequences per treatment replicate was performed to standardize comparisons. (A) Major OTUs identified in jejunal contents. (B) Major OTUs identified in cecal contents.

CPN23 for OTU_01 and OTU_02, respectively. Other abundant bacterial genera observed in jejunal contents were: *Staphylococcus*, *Enterococcus*, *Jeotgalicoccus*, *Corynebacterium*, and *Streptococcus* (Table A1.2).

The most abundant bacterial phyla identified in cecal contents were Firmicutes, Tenericutes, Bacteroidetes, and Proteobacteria. Firmicutes were consistently found at a higher relative abundance than any other phylum in the cecum, with relative abundances among the treatment replicates ranging from 87.82% (control rep 4) to 92.86% (virginiamycin rep 2). Relative abundances among treatment replicates for other predominant phyla identified in cecal contents included the Tenericutes (mean = 1.43%), Bacteroidetes (mean = 1.28%), and Proteobacteria (mean = 0.84%). In the broiler cecum, OTUs that individually were less than one percent of the total accounted for on average 68.22% of the total bacterial community when combined (Figure A1.2B). In particular, OTU_04, classified as belonging to the genus *Faecalibacterium*, was the most abundant bacterial group in cecal contents for every treatment replicate, with one exception (BMD replicate 2), ranging from 8.25% (control rep 3) to 20.41% (virginiamycin rep 1) of the total cecal sequences observed within a single dietary treatment replicate. OTUs classified as *Lactobacillus* were greatly reduced in abundance in cecal contents compared to jejunal contents. However, the two most abundant *Lactobacillus* OTUs in the jejunum, OTU_01 and OTU_02, were also among the 11 most abundant OTUs in the cecum. The average percent relative abundance for OTU_01 in cecal contents across all treatment replicates was 4.69%, while the average for OTU_02 was 1.08%. Other abundant genera consistently observed in cecal contents included: *Oscillospira*, *Ruminococcus*, *Blautia*, *Coprococcus*, *Streptococcus*, *Sutterella*, *Coprobacillus*, and *Enterococcus* (Table A1.3).

Table A1.2: Mean percent relative abundances of major genera in jejunal contents

Genus	Diet Group		
	BMD	Control	Virginiamycin
Lactobacillus	58.07 ± 5.15	75.96 ± 15.26	73.40 ± 12.21
Staphylococcus	2.90 ± 1.98	1.58 ± 1.42	1.61 ± 2.37
Enterococcus	1.43 ± 0.46	1.33 ± 1.59	1.75 ± 1.37
Jeotgalicoccus	1.05 ± 1.21	0.79 ± 0.56	0.43 ± 0.73
Corynebacterium	0.80 ± 0.52	1.01 ± 0.97	0.24 ± 0.27
Streptococcus	0.61 ± 0.71	0.67 ± 0.81	0.27 ± 0.24
Salinicoccus	0.36 ± 0.29	0.90 ± 0.57	0.09 ± 0.09
Yaniella	0.37 ± 0.29	0.76 ± 0.79	0.09 ± 0.06
Turicibacter	0.21 ± 0.36	0.12 ± 0.17	0.68 ± 1.32
Facklamia	0.14 ± 0.19	0.25 ± 0.12	0.08 ± 0.13
Weissella	0.24 ± 0.27	0.06 ± 0.08	0.10 ± 0.06
Others	33.81 ± 5.69	16.57 ± 10.57	21.27 ± 10.08

Mean ± standard deviation for four trials

Impact of growth-promoting antibiotics on the GI bacterial community

Feeding virginiamycin resulted in a less diverse bacterial community in broiler intestines (Figure A1.3), relative to the BMD-fed group. In particular, evidence of a dietary group effect on cecal bacterial community diversity ($P = 0.021$), as determined using Shannon's index, was observed. Pairwise comparisons of the dietary treatments revealed a significant difference in the least squares (LS) means (control = 5.74, virginiamycin = 5.44, and BMD = 5.65, standard error = 0.06) of the control and virginiamycin treatments ($P = 0.023$), but not between the virginiamycin and BMD treatments ($P = 0.138$), or the control and BMD treatments ($P = 0.881$). 454 pyrosequencing also revealed differences among the diets in the abundance of specific bacterial taxa (Table A1.4, Figure A1.4). Specifically, a significant dietary treatment effect on the percent relative abundance of OTU_02 (genus *Lactobacillus*) was observed in both the jejunum ($P = 0.004$) and cecum ($P = 0.026$). Feeding virginiamycin resulted in an enrichment of

Table A1.3: Mean percent relative abundances of major genera in cecal contents

Genus	Diet Group		
	BMD	Control	Virginiamycin
Faecalibacterium	13.54 ± 1.70[B]	14.99 ± 1.20[B]	21.22 ± 2.64[A]
Oscillospira	11.33 ± 1.44	9.96 ± 2.80	9.12 ± 0.57
Lactobacillus	10.14 ± 3.64	7.24 ± 1.81	6.45 ± 1.88
Ruminococcus	4.65 ± 0.87	4.49 ± 0.33	4.23 ± 0.09
Blautia	4.16 ± 0.43	3.88 ± 0.44	4.45 ± 0.72
Coprococcus	0.95 ± 0.29	2.12 ± 2.39	1.24 ± 0.26
cc_115	1.47 ± 0.49	1.46 ± 0.51	0.78 ± 0.18
02d06	0.67 ± 0.24	0.62 ± 0.49	1.14 ± 0.40
Streptococcus	0.50 ± 0.75	1.43 ± 1.85	0.00 ± 0.00
Sutterella	0.53 ± 0.19	0.52 ± 0.14	0.64 ± 0.10
Coprobacillus	0.46 ± 0.03	0.36 ± 0.20	0.50 ± 0.09
Enterococcus	0.44 ± 0.15	0.33 ± 0.32	0.22 ± 0.09
AF12	0.32 ± 0.14	0.35 ± 0.23	0.28 ± 0.07
Bacteroides	0.18 ± 0.20	0.35 ± 0.31	0.26 ± 0.23
Dehalobacterium	0.16 ± 0.04	0.14 ± 0.05	0.19 ± 0.04
Salmonella	0.10 ± 0.08	0.20 ± 0.16	0.14 ± 0.04
Others	50.39 ± 3.63	51.57 ± 3.20	49.14 ± 2.84

Mean ± standard deviation for four trials
[A-B]Different letters indicate highly significant differences among the means in the same row ($P < 0.01$).

these bacteria regardless of GI section. In the jejunum, the difference in the LS mean percent relative abundance of OTU_02 (control = 15.09, virginiamycin = 33.14, and BMD = 7.14, standard error = 3.99) between the virginiamycin and BMD treatments was highly significant ($P = 0.004$) while the difference between the virginiamycin and control treatments was merely significant ($P = 0.033$). In the cecum, only the difference in the LS mean percent relative abundance of OTU_02 (control = 0.90, virginiamycin = 1.83, and BMD = 0.51, standard error = 0.28) between the virginiamycin and BMD treatments was observed to be significant ($P = 0.029$). Feeding virginiamycin to broiler chickens also enriched for the genus *Faecalibacterium*

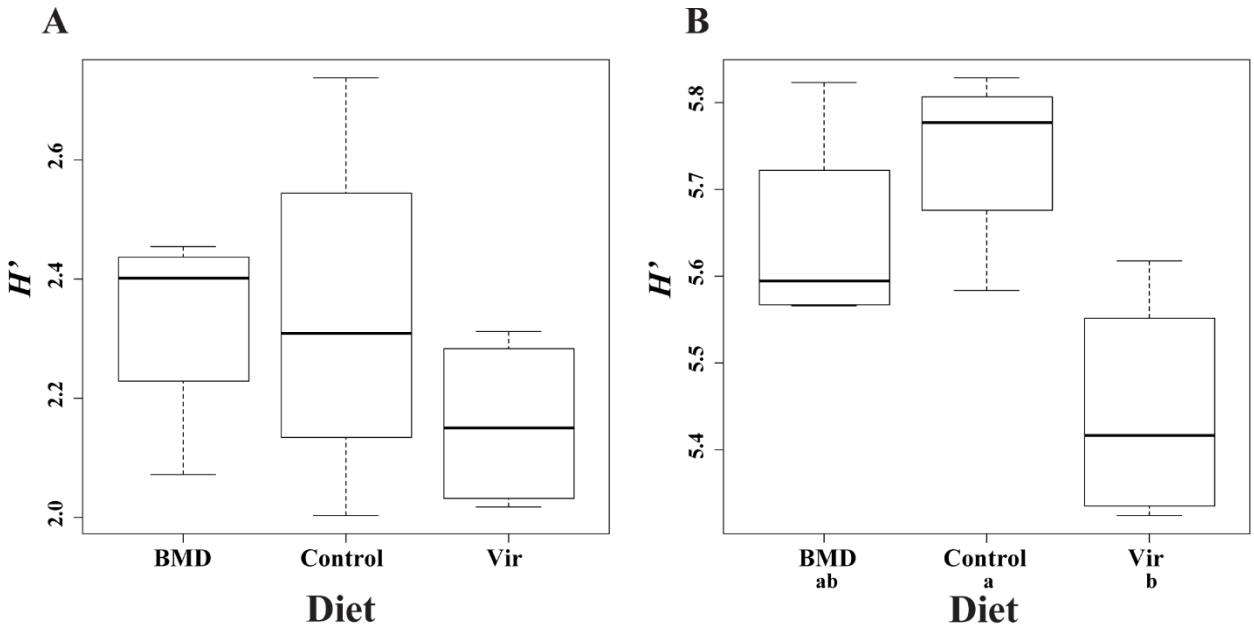


Figure A1.3: Diversity of bacterial communities in broiler intestines. Shannon's diversity (*H'*) indices for bacterial communities in (A) jejunal contents and (B) cecal contents for each dietary treatment are represented by box plots. Each individual box plot consists of means from 4 replicates, each calculated from 10 subsampled broilers. The heavy black bars indicate the median value. The bottom of the box and top of the box respectively represent the first and third quartiles. The dashed lines extend to the minimum and maximum values or 1.5 times the interquartile range, whichever is larger. Letters below the treatment label (a-b) indicate significant differences ($P < 0.05$) among the group means.

in the cecum compared to both the BMD diet ($P = 0.001$) and control diet ($P = 0.004$). Specifically, the percent relative abundance of OTU_04 (genus *Faecalibacterium*), the most abundant bacterial group in the cecum, was enriched by feeding virginiamycin over the BMD treatment ($P = 0.028$) and the control treatment ($P = 0.036$).

Clostridium perfringens in the GI microbiota of broiler chickens

We used qPCR to specifically assess the abundance of *C. perfringens* in jejunal and cecal contents. The relative abundance of *C. perfringens* in jejunal contents, based upon the ratio of *C. perfringens* 16S rRNA to total Bacterial 16S rRNA, was most typically between 1:100,000

Table A1.4: Major bacterial groups with notable dietary treatment effects

Group	Classification[1]	% Relative Abundance[2]			Test	P-value
		Control	Vir[3]	BMD		
OTU_01: Cecum	*Lactobacillus* (100)	4.83	2.63	7.06	Kruskal-Wallis	0.0592
OTU_02: Jejunum	*Lactobacillus* (100)	15.01^{b}	33.14^{Aa}	7.14^{B}	ANOVA	0.0037
OTU_02: Cecum	*Lactobacillus* (100)	0.90^{ab}	1.83^{a}	0.51^{b}	ANOVA	0.0255
OTU_04: Cecum	*Faecalibacterium* (100)	11.19^{b}	16.79^{a}	10.90^{b}	ANOVA	0.0150
Faecalibacterium: Cecum[3]	Ruminococcaceae	14.99^{B}	21.22^{A}	13.54^{B}	ANOVA	0.0008

[1]Most resolved phylogenetic match using the Greengenes 16S rRNA gene database with consensus confidence/bootstrap value in parentheses
[2]Abundances for groups tested using the rank-based Kruskal-Wallis test represent the median of 4 trails value while counts for groups tested using ANOVA represent the mean of 4 trials
[3]The *Faecalibacterium* group represents all of the OTUs classified as genus *Faecalibacterium* and family Ruminococcaceae combined
Vir = Virginiamycin
[A-B]Different letters indicate highly significant differences among the means in the same row ($P < 0.01$)
[a-b]Different letters indicate significant differences among the means in the same row ($P < 0.05$)

(log10 = -5) and 1:1,000,000 (log10 = -6) gene copies (Figure A1.5A). The most common result when testing cecal samples was no detection (Figure A1.5B). A considerable range of at least six orders of magnitude in *C. perfringens* levels among the different broilers sampled was observed for both jejunal and cecal samples. Table A1.5 shows the median log10 relative abundances of *C. perfringens* observed for each of the dietary treatment replicates. The highest median log10 relative abundance of *C. perfringens* in jejuna for any one dietary treatment replicate was -3.97 (control rep 3). The highest median log10 relative abundance of *C. perfringens* in ceca for any one dietary treatment replicate was -3.42 (virginiamycin rep 3). The lowest median log10 relative abundance of *C. perfringens* in the jejunum for any one dietary treatment replicate was -6.19 (control rep 1). Several dietary treatment replicates had median cecal log10 relative abundances that were below the level of detection: control rep 1, virginiamycin rep 4, BMD rep 1, and BMD rep 2. No evidence for a difference among the diets

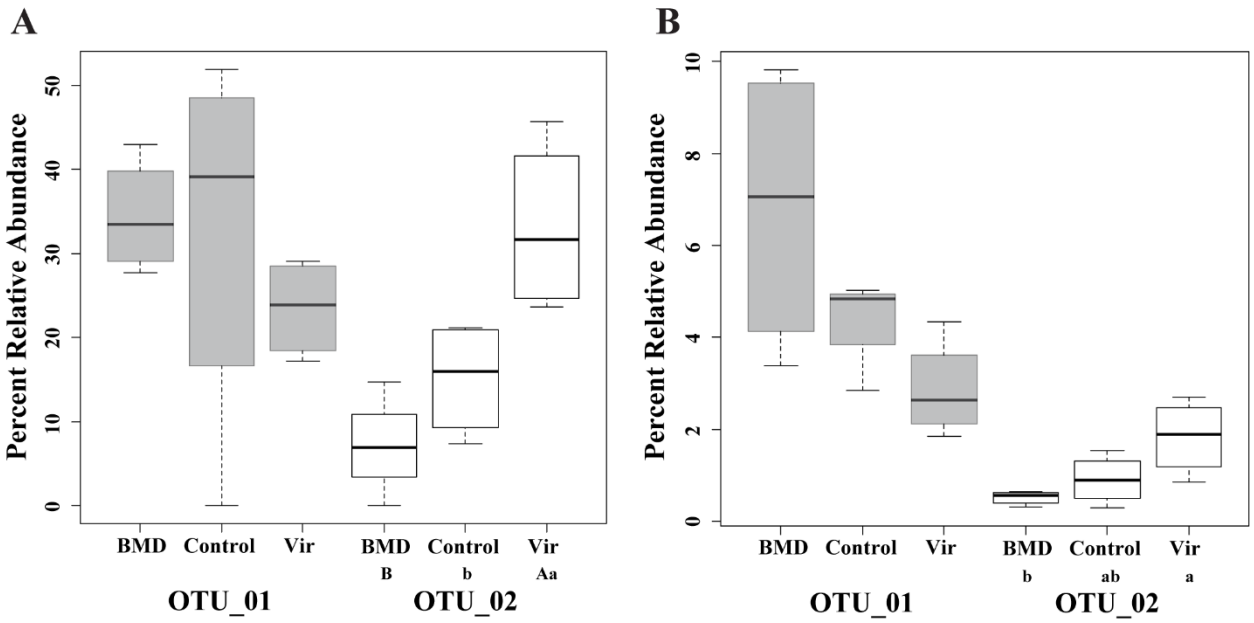


Figure A1.4: Dietary treatment effects on two distinct intestinal *Lactobacillus* populations. Box plots represent the distributions of the observed percent relative abundance of OTU_01 (shaded) and OTU_02 (unshaded) for each dietary treatment in (A) jejunal contents and (B) cecal contents. Each individual box plot consists of means from 4 replicates, each calculated from 10 subsampled broilers. The heavy black bars indicate the median value. The bottom of the box and top of the box respectively represent the first and third quartiles. The dashed lines extend to the minimum and maximum values or 1.5 times the interquartile range, whichever is larger. Different letters, A-B, indicate highly significant differences among the treatment means ($P < 0.01$). Letters below the treatment label (a-b) indicate significant differences among the group means ($P < 0.05$).

was observed for levels of *C. perfringens* in jejunal contents ($P = 0.569$), or cecal contents ($P = 0.675$). In all cases, levels of *C. perfringens* were numerically highest during the third replicate grow-out.

A1.4 Discussion

The broiler chicken GI tract is colonized by a dense community of microorganisms that is intimately connected to the overall heath and development of its animal host. In livestock, there is considerable interest in understanding how this community contributes to the efficient conversion of feed into growth. Improvements in feed conversion associated with dietary

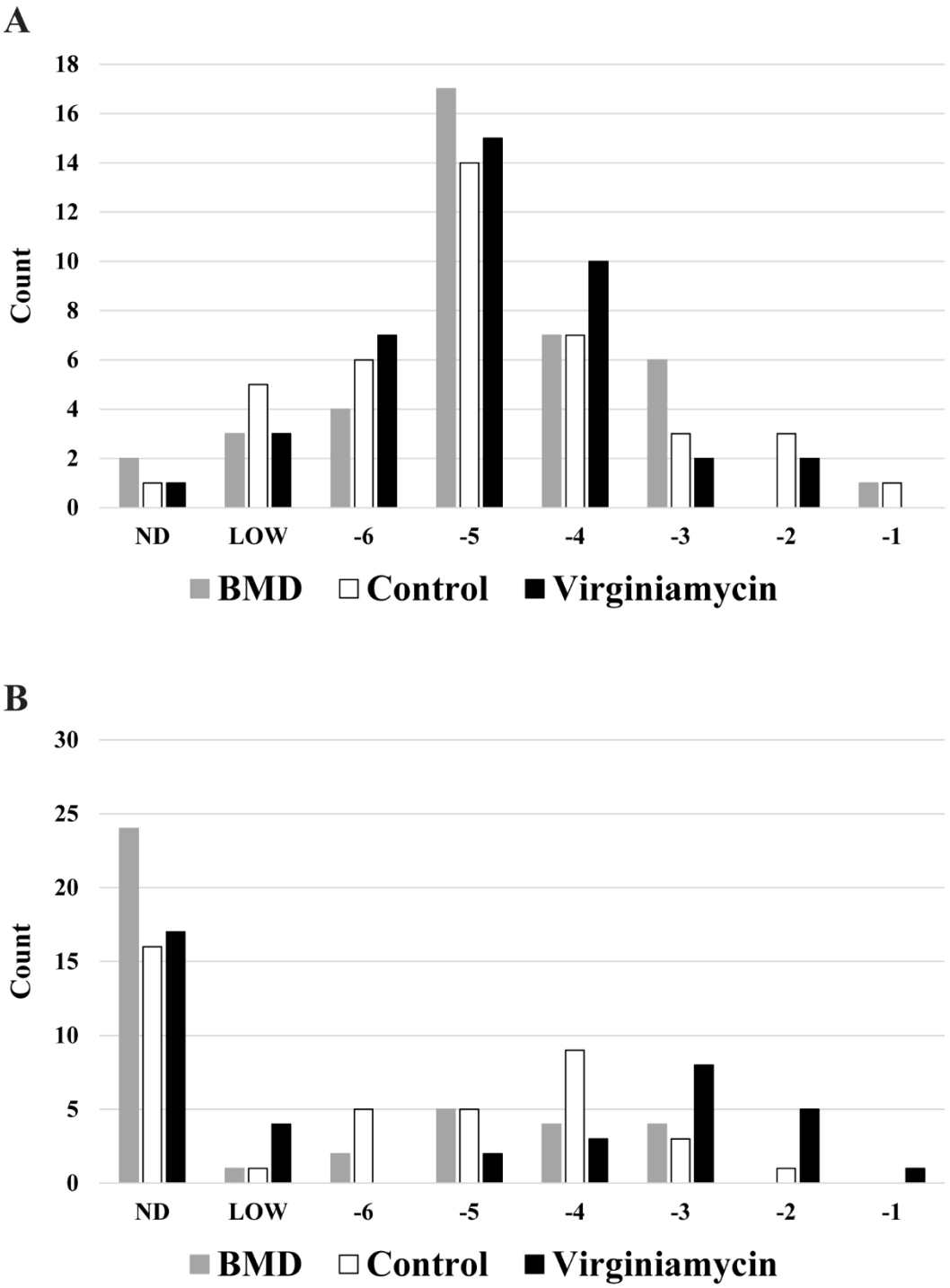


Figure A1.5: Distributions of log10 relative levels of *C. perfringens*. Quantitative PCR (qPCR) was used to determine the relative abundance of *C. perfringens* in broiler (A) jejunal contents and (B) cecal contents. Counts (y-axis) represents the number of samples falling within a particular range of abundances (x-axis). Open bars represent the control diet, black bars the virginiamycin diet, and grey bars the BMD diet. Samples near or below the assay limit of detection were categorized as LOW or ND = not detected.

Table A1.5: Median log10 relative abundance of *Clostridium perfringens*

Gut Section	Diet Group	Log10 level[1]			
		Rep 1	Rep 2	Rep 3	Rep 4
Jejunum	Control	-6.19	-5.50	-3.97	-5.85
Jejunum	Virginiamycin	-5.50	-5.44	-4.29	-6.15
Jejunum	BMD	-5.32	-5.65	-3.99	-5.35
Cecum	Control	ND	-5.68	-4.44	-6.09
Cecum	Virginiamycin	-6.17	-6.32	-3.42	ND
Cecum	BMD	ND	ND	-4.43	-6.48

[1]Median log10 relative abundance of *C. perfringens* 16S rRNA gene copies per total bacterial 16S rRNA gene copies from the 10 birds sampled per trial
ND = no *C. perfringens* specific signal detected in any of the three qPCR technical replicates

supplementation with antibiotics are thought to involve GI tract microbial communities, but this connection remains poorly understood (1). Here, we sought to elucidate the effects of virginiamycin and BMD on the broiler GI tract bacterial community under simulated commercial production conditions. Four replicate grow-outs consisting of three pens, one for each different dietary condition, were performed consecutively with diets for each pen remaining constant. Overall, 120 total birds, ten from each pen, were sampled in order to account for the considerable inter-individual variation in the GI microbiota previously reported for these animals (21).

The jejunum and cecum of broiler chickens serve distinct roles in digestion and harbor different microbial communities (17, 19, 22, 23). The jejunum is a site of enzymatic degradation and absorption of proteins, carbohydrates, and fats, while the longer retention times for digesta in the cecum allow for more extensive fermentation of recalcitrant plant polysaccharides by resident microbes (18, 19). In jejunal contents we observed a bacterial community that was largely dominated by populations of *Lactobacillus*. Two phylotypes in particular, OTU_01 and OTU_02, which are closely related to *L. crispatus* and *L. johnsonii*, respectively, typically accounted for approximately 50% of the total bacterial community in jejunal contents.

Lactobacillus are facultative anaerobes that thrive in nutritionally rich environments as well as on the mucosal surfaces of animals (24). The average whole gut passage time for the chicken is believed to be 3.5 hours, providing a strong selective pressure in the jejunum for microorganisms capable of colonizing the mucosa and/or exhibiting particularly fast growth rates under nutrient rich conditions, such as members of the *Lactobacillus* (25).

In contrast to the bacterial community observed in the jejunum, the cecal bacterial communities were substantially more diverse and dominated by a community of Clostridia. The most abundant phylotype observed in cecal contents was consistently OTU_04, which was classified as a member of the genus *Faecalibacterium*. Conditions in the broiler cecum are completely anaerobic and there is little soluble carbohydrate compared with the more proximal sections of the GI tract (18, 19). This suggests that, like other Clostridiales, OTU_04 represents a group of bacteria that are strictly anaerobic and potentially capable of subsisting on plant cell wall polysaccharides (26).

One of the goals of this study was to test the hypothesis that feeding broiler chickens virginiamycin has a more pronounced effect on the GI tract microbiota than feeding BMD (9). Our results support this, as we observed that virginiamycin significantly reduced diversity in the cecum compared to the control diet, whereas BMD did not. No evidence for an antibiotic effect on community diversity was observed in the jejunum. In addition to the changes observed in overall cecal community diversity, an enrichment for OTU_02 (genus *Lactobacillus*) compared to the control diet in the jejunum and an enrichment for the genus *Faecalibacterium* in the cecum was also observed in broilers receiving the virginiamycin supplemented diet. An enrichment in cecal Clostridia, a group that includes *Faecalibacterium*, associated with virginiamycin supplementation has been described previously (6), and may explain some of the increases in

feed efficiency that are typically observed for this feeding regimen as related organisms have been recently correlated with improved dietary energy extraction in broilers (27). For example *Faecalibacterium prausnitzii* has been shown to produce butyrate, a fermentation product demonstrated to be beneficial to intestinal epithelial cells (28). In contrast, BMD was not associated with any significant alterations in the dominant bacterial groups present in the jejunum or cecum. Despite the lack of statistical evidence, BMD did numerically enrich for OTU_01 (genus *Lactobacillus*), compared to virginiamycin, in both the jejunum and cecum. In the cecum, this difference was observed to be approaching statistical significance.

Although both virginiamycin and BMD have been demonstrated to be effective in controlling NE caused by *C. perfringens* (14, 15), differences in how these compounds affect *C. perfringens* levels in the poultry environment have been reported (9). We attempted to model our experiment after the aforementioned study in order to determine if a similar result was observed for levels of this pathogen in the GI tract. However, no evidence of a dietary group effect was observed on *C. perfringens* levels in either the jejunum or cecum. It is important to note that no evidence of NE, or other pathology associated with *C. perfringens*, was observed during these four replicate grow-outs. This could potentially explain why we observed no difference in *C. perfringens* levels, as it is difficult to reproduce the exact microbial challenge experienced between the studies. Nevertheless, broilers from the BMD dietary treatment did tend to have numerically lower levels of *C. perfringens* in the cecum, relative to the other two dietary treatments over the course of the first three replicate grow-outs. Also notable is the observation that for every dietary treatment, the highest levels of *C. perfringens* were observed for the third replicate grow-out. These results suggest that the time in between the grow-outs may be an important determinant in the levels of *C. perfringens* experienced by a flock. During

the first three grow-outs of this study, cycle downtimes were similar to industry standards (~2 weeks); however, the downtime between grow-outs 3 and 4 (~4 weeks) was considerably longer than is typical during commercial production. Future work should test the hypothesis that extended downtimes allow for enough outgrowth of environmental organisms to diminish *C. perfringens* levels in the litter which subsequently leads to lower levels in the GI tract.

In conclusion, this study represents the most resolved characterization of bacterial communities in the jejunum and cecum of market-aged broiler chickens raised under simulated commercial production conditions receiving diets supplemented with either BMD or virginiamycin. Our results demonstrate that virginiamycin has a relatively broader impact on broiler GI tract bacterial communities, as compared to BMD. Specifically, we observed enrichments in a group of *Lactobacillus*, closely related to *L. johnsonii*, in the jejunum and cecum as well as an enrichment in *Faecalibacterium* in the cecum accompanying virginiamycin supplementation. Although no significant differences in the abundance of *C. perfringens* were observed among the dietary groups, a potential connection between grow-out cycle downtime and the GI tract bacterial community composition in 6 week old broilers was identified that warrants further study.

A1.5 Materials & Methods

Animal treatments

Four consecutive 44-day simulated commercial grow-outs were conducted at Colorado Quality Research, Inc. (CQR) (Wellington, CO) between September 2012 and May 2013. The days between the respective replicate grow-outs (downtimes) were as follows: rep 1 to 2 = 11 days, rep 2 to 3 = 19 days, rep 3 to 4 = 27 days. At the beginning of each grow-out (day = 0),

540 1-day old straight-run (1:1 male:female) Cobb 500 broiler chickens were randomly allocated among three experimental groups (180 birds per group). Each experimental group was maintained in individual floor pens within one environmentally controlled building. A solid, plastic, divider approximately 48 inches in height between each pen was employed to avoid cross-contamination among the groups. Prior to chick placement for the first grow-out, used litter from a previous study at CQR (unpublished) in which no antibiotics were used was thoroughly mixed and placed into all three floor pens. For all subsequent grow-outs, associations between each pen and its assigned experimental group were strictly maintained. Chickens were fed basal starter (days 0 – 17), grower (days 17 – 30) and finisher (days 30 – 44) diets manufactured at the CQR feed mill using standard CQR formulated corn-soy diets representative of commercial broiler production diets (Agristats Middle 50%). Experimental groups consisted of these diets supplemented with either: 50 grams per ton bacitracin methylene disalicylate (BMD 50, Zoetis, Florham Park, NJ), 20 grams per ton virginiamycin (Stafac, Phibro Animal Health, Teaneck, NJ), or un-supplemented (control). Feed and water were provided *ad libitum* throughout each grow-out period. Chickens received Vaxxitek (Merial, Duluth, GA) vaccine for Marek's HVT + IBDV and Inovocox EM-1 (Zoetis, Florham Park, NJ) vaccine for Coccidia in ovo. On day 0 birds were vaccinated against Newcastle disease and infectious bronchitis with Poulvac Aero (Zoetis, Florham Park, NJ) by spray application.

Sample collection

At day 44 of each grow-out, 10 broilers were randomly selected from each of the 3 experimental groups and humanely euthanized via cervical dislocation. GI tracts, from the duodenal loop to just past the ileocecal junction, were removed from each chicken prior to

further dissection. Intestinal contents were collected from the jejunum as follows: a 6 inch section of the intestines surrounding the Meckel's diverticulum was excised, contents from this section gently squeezed out, and collected into a sterile 50 mL conical test tube. The contents were weighed, diluted 5x in sterile $T_{50}E_{50}$ + sucrose buffer (50 mM Tris-HCl, 50 mM EDTA, 15% sucrose, pH = 8.0), homogenized, and frozen at -20° C. Similarly, the tips of the cecal lobes were cut off, the contents squeezed out, diluted, and stored at -20° C. In total, 240 samples (two intestinal sections multiplied by 30 birds multiplied by 4 replicate grow-outs) were collected and subjected to further analysis by pyrosequencing and quantitative PCR.

DNA extraction

DNA extraction was performed on 1 mL of diluted intestinal contents. Cells were harvested from the intestinal contents by centrifugation and re-suspended in 1 mL DNA extraction buffer (100 mM Tris-HCl, 10 mM EDTA, 0.15 M NaCl, pH = 8.0) containing 5 mg mL^{-1} lysozyme. After incubation at 39° C for 15 min to digest bacterial cell walls the cell suspension was transferred to a 2 mL microcentrifuge tube containing 0.5 g 0.1 mm zirconium beads, 700 µL equilibrated phenol (pH = 8.0), and 50 µL 20% SDS. The samples then underwent bead-beating for 2 min followed by incubation at 60° C for 10 min and a second round of bead-beating for 2 min. The organic and aqueous phases were separated by centrifugation and aqueous phase collected into a fresh 1.5 mL microcentrifuge tube. Four additional extractions of the aqueous phase were performed, one with equilibrated phenol, two with phenol:chloroform:isoamyl alcohol (25:24:1), and one with chloroform. The DNA was recovered from solution by precipitation with 3 M Na acetate and isopropanol, and re-suspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH = 8.0).

16S library preparation and pyrosequencing

PCR amplification of the variable V6-V8 region of the 16S rRNA gene for amplicon pyrosequencing was achieved using primers containing the Roche 454 A or B Titanium sequencing adaptors on the forward primer: 926F-5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGAAACTYAAAKGAATTGACGG-3'. The reverse primer included one of 30 barcodes, 5 bp in length, as indicated by XXXXX: 1392R-5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-XXXXX-ACGGGCGGTGTGTRC-3'. Five 20 µL reactions including: 2 ng template DNA, 0.5 µM of each primer, and 17 µL high-fidelity DNA polymerase Platinum Blue master mix (Invitrogen by Life Technologies, Grand Island, NY) were performed for each sample. PCR conditions consisted of an initial denaturation at 94° C for 2 min followed by 30 cycles of 94° C for 30 sec, 50° C for 45 sec, and 68° C for 1 min 45 sec, with a final extension at 68° C for 10 min. Products from the five reactions were pooled, purified, and concentrated to 40 µL using the PureLink® Quick PCR Purification Kit (Invitrogen by Life Technologies). Cleaned products were purified further by excising the appropriately sized amplicon (~550 bp) from a 1% low-melt agarose gel after electrophoresis, followed by recovery using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp., Irvine, CA). Each sample was quantified using a Qubit® Fluorometer (Invitrogen, San Diego, CA), and pooled to create a single library at 1×10^9 molecules μL^{-1}. The library was subsequently diluted in ultrapure H_2O to a concentration of 1×10^6 molecules μL^{-1} and used as template for emPCR at a ratio of 0.8 molecules per DNA capture bead. Bead recovery and sequencing were performed following the manufacturer's guidelines for a Roche 454 GS Junior with the Lib-L kit and Titanium chemistry (Roche Applied Science, Indianapolis, IN).

Quantitative real-time PCR

Two distinct probe-based quantitative real-time PCR (qPCR) assays were employed to quantify the abundance of total bacteria (29) and *Clostridium perfringens* (10), respectively, in each sample of intestinal contents in order to estimate their relative abundance. Each 25 µL general Bacteria reaction consisted of 12.5 µL Premix Ex Taq (Clontech, Mountain View, CA), 0.4 µM forward primer (5'-ACTCCTACGGGAGGCAG-3'), 0.4 µM reverse primer (5'-GACTACCAGGGTATCTAATCC-3'), 0.2 µM FAM-labeled probe (5'-TGCCAGCAGCCGCGGTAATAC-3'), 8.5 µL sterile H_2O, and DNA template (1 ng μL^{-1} and 10 ng μL^{-1} for jejunum samples and cecum samples respectively). Each 25 µL *C. perfringens*-specific reaction contained 12.5 µL Premix Ex Taq (Clontech), 0.6 µM forward primer (5'-CGCATAACGTTGAAAGATGG-3'), 0.6 µM reverse primer (5'-CCTTGGTAGGCCGTTACCC-3'), 0.3 µM FAM-labeled probe (5'-TCATCATTCAACCAAAGGAGCAATCC-3'), 8.5 µL sterile H_2O, and DNA template (1 ng μL^{-1} and 10 ng μL^{-1} for jejunum samples and cecum samples respectively). The same cycling conditions were used for both assays and are as follows: initial denaturation at 95° C for 30 sec followed by 45 cycles of 95° C for 5 sec and 56° C for 30 sec. A plate read step was performed after the annealing/extension step of each cycle. Real-time PCR was performed using a Bio-Rad CFX96 Real-time system (Bio-Rad, Hercules, CA). Each assay for each sample was performed in triplicate and compared to a standard curve, consisting of serially diluted *C. perfringens* genomic DNA, included in each thermocycler run to determine the amount of target in the sample. Relative abundance of *C. perfringens* in a sample was determined by dividing the amount of *C. perfringens* target observed for a given sample by the amount of Bacterial target

observed. *C. perfringens* mean log10 abundance and standard deviation was calculated for each sample from the three qPCR technical replicates. Samples for which one or two of the three technical replicates failed to amplify were typically at or near the level of detection for the assay and were categorized as LOW. Samples for which none of the three technical replicates crossed the cycle threshold were categorized as Not Detected (ND). A small number of samples exhibited considerable variation (log10 standard deviation > 0.5) among the three technical replicates. These samples were also typically near the limit of detection for the assay and were therefore categorized as LOW.

Processing and analysis of pyrosequencing data

Sequence processing and data analysis was performed using the program mothur v.1.13.3 (30) with default command parameters, unless specified. Flowgrams were trimmed (min = 360, max = 720) and de-noised using the *shhh.flows* command, the mothur implementation of the AmpliconNoise algorithm (31). The resulting sequences were trimmed (pdiffs = 1, bdiffs = 0, maxhomop = 6, minlength = 250) followed by alignment against the SILVA 16S rRNA gene reference alignment database (32). Chimera detection was performed using the *chimera.uchime* command (http://drive5.com/uchime) on a screened version of the alignment (*filter.seqs*) that had been reduced using *unique.seqs* and *pre.cluster* (diffs = 2). Sequences classified as either domain Eukaryota or domain Archaea were removed during sequence processing and were therefore excluded from all subsequent analyses. The *cluster.split* command was used to assign sequences to operational taxonomic units (OTUs) using the furthest neighbor algorithm with a cutoff of 0.03, considered to correspond to species level identification in bacteria (33). Classification of OTUs was achieved using the Greengenes database (34), with a consensus

confidence/bootstrap value threshold of at least 80 percent. Mean diversity and coverage metrics were obtained using the command *summary.single* (calc = sobs-shannon-coverage, label = 0.03, subsample = T): number of observed OTUs, Shannon diversity, and Good's coverage, which performs 1,000 iterations to estimate the respective means for each group. Analysis of community similarities (ANOSIM) (35) and principal coordinates analysis (PCoA) was performed using the *anosim* and *pcoa* commands, respectively. Representative DNA sequences for OTUs of interest were obtained using the *get.oturep* command, and were analyzed by BLAST search (http://www.ncbi.nlm.nih.gov/BLAST/) to determine the closest matches in the GenBank repository. All bacterial community, or individual taxa, comparisons were normalized by subsampling to the minimum number of processed sequences recovered for a single treatment replicate (50,408 sequences).

Statistical analyses

The unit of replication for all statistical tests was the individual pen with the ten birds sampled per pen treated as subsamples. In total, there were four pens per dietary treatment, one from each replicate grow-out. The distributions of all data to be evaluated were examined individually by constructing density histograms prior to implementation of any particular statistical test. Data with a distribution that was roughly symmetrical and bell-shaped, or approximately normally distributed, was evaluated for constant variance among the dietary treatments using Levene's test in R (version 3.0.0). If Levene's test failed to reject the null hypothesis (alpha = 0.05) of equal variance among treatments, the dependent variable of interest was tested using one-way ANOVA in SAS (version 9.4). If the overall P-value was significant at alpha = 0.05, pairwise comparisons were performed. Bonferroni corrected P-values were used

for all pairwise comparisons (alpha = 0.05). Data exhibiting distributions that were highly skewed, obviously not normal, or categorical (*e.g.* the *C. perfringens* qPCR data), were evaluated with the nonparametric Kruskal-Wallis rank sum test in R (alpha = 0.05).

Sequence accession numbers

Standard flowgram format (sff) files for each of the 240 samples analyzed by 454 pyrosequencing have been deposited at the National Center for Biotechnology Information (NCBI) Short Read Archive (http://www.ncbi.nlm.nih.gov/sra) and can be found under project accession PRJNA260213.

A1.6 References

1. Dibner JJ, Richards JD. 2005. Antibiotic growth promoters in agriculture: history and mode of action. Poult Sci 84:634–643.
2. Kim HB, Borewicz K, White BA, Singer RS, Sreevatsan S, Tu ZJ, Isaacson RE. 2012. Microbial shifts in the swine distal gut in response to the treatment with antimicrobial growth promoter, tylosin. Proc Natl Acad Sci USA 109:15485–15490.
3. Coates ME, Fuller R, Harrison GF, Lev M, Suffolk SF. 1963. A comparison of the growth of chicks in the Gustafsson germ-free apparatus and in a conventional environment, with and without dietary supplements of penicillin. Br J Nutr 17:141–150.
4. Butaye P, Devriese LA, Haesebrouck F. 2003. Antimicrobial growth promoters used in animal feed: effects of less well known antibiotics on gram-positive bacteria. Clin Microbiol Rev 16:175–188.
5. Dumonceaux TJ, Hill JE, Hemmingsen SM, Van Kessel AG. 2006. Characterization of intestinal microbiota and response to dietary virginiamycin supplementation in the broiler chicken. Appl Environ Microbiol 72:2815–2823.
6. Lu J, Hofacre C, Smith F, Lee MD. 2008. Effects of feed additives on the development on the ileal bacterial community of the broiler chicken. Animal 2:669–676.
7. Danzeisen JL, Kim HB, Isaacson RE, Tu ZJ, Johnson TJ. 2011. Modulations of the chicken cecal microbiome and metagenome in response to anticoccidial and growth promoter treatment. PLoS ONE 6:e27949.
8. Torok VA, Allison GE, Percy NJ, Ophel-Keller K, Hughes RJ. 2011. Influence of antimicrobial feed additives on broiler commensal posthatch gut microbiota development and performance. Appl Environ Microbiol 77:3380–3390.
9. LaVorgna M, Schaeffer JL, Bade D, Dickson J, Cookson K, Davis SW. 2013. Performance of broilers fed a broader spectrum antibiotic (virginiamycin) or a narrower spectrum antibiotic (bacitracin methylene disalicylate) over 3 consecutive grow-out cycles. J Appl Poult Res 22:574–582.
10. Wise MG, Siragusa GR. 2005. Quantitative detection of *Clostridium perfringens* in the broiler fowl gastrointestinal tract by real-time PCR. Appl Environ Microbiol 71:3911–3916.
11. Stanley D, Keyburn AL, Denman SE, Moore RJ. 2012. Changes in the caecal microflora of chickens following *Clostridium perfringens* challenge to induce necrotic enteritis. Vet Microbiol 159:155–162.
12. Keyburn AL, Boyce JD, Vaz P, Bannam TL, Ford ME, Parker D, Di Rubbo A, Rood JI, Moore RJ. 2008. NetB, a new toxin that is associated with avian necrotic enteritis caused by *Clostridium perfringens*. PLoS Pathog 4:e26.
13. Prescott JF. 1979. The prevention of experimentally induced necrotic enteritis in chickens by avoparcin. Avian Dis 23:1072–1074.
14. Wicker DL, Iscrigg WN, Trammell JH. 1977. The control and prevention of necrotic enteritis in broilers with zinc bacitracin. Poult Sci 56:1229–1231.
15. George BA, Quarles CL, Fagerberg DJ. 1982. Virginiamycin effects on controlling necrotic enteritis infection in chickens. Poult Sci 61:447–450.
16. Stutz MW, Johnson SL, Judith FR. 1983. Effects of diet and bacitracin on growth, feed efficiency, and populations of *Clostridium perfringens* in the intestine of broiler chicks. Poult Sci 62:1619–1625.

17. Yegani M, Korver DR. 2008. Factors affecting intestinal health in poultry. Poult Sci 87:2052–2063.
18. Yeoman CJ, Chia N, Jeraldo P, Sipos M, Goldenfeld ND, White BA. 2012. The microbiome of the chicken gastrointestinal tract. Anim Health Res Rev 13:89–99.
19. Oakley BB, Lillehoj HS, Kogut MH, Kim WK, Maurer JJ, Pedroso A, Lee MD, Collett SR, Johnson TJ, Cox NA. 2014. The chicken gastrointestinal microbiome. FEMS Microbiol Lett 360:100–112.
20. Wei S, Morrison M, Yu Z. 2013. Bacterial census of poultry intestinal microbiome. Poult Sci 92:671–683.
21. Stanley D, Geier MS, Hughes RJ, Denman SE, Moore RJ. 2013. Highly variable microbiota development in the chicken gastrointestinal tract. PLoS ONE 8:e84290.
22. Gong J, Si W, Forster RJ, Huang R, Yu H, Yin Y, Yang C, Han Y. 2007. 16S rRNA gene-based analysis of mucosa-associated bacterial community and phylogeny in the chicken gastrointestinal tracts: from crops to ceca. FEMS Microbiol Ecol 59:147–157.
23. Choi JH, Kim GB, Cha CJ. 2014. Spatial heterogeneity and stability of bacterial community in the gastrointestinal tracts of broiler chickens. Poult Sci 93:1942–1950.
24. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Benson A, Baldwin K, Lee J-H, Díaz-Muñiz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O'Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D. 2006. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA 103:15611–15616.
25. Pan D, Yu Z. 2014. Intestinal microbiome of poultry and its interaction with host and diet. Gut Microbes 5:108–119.
26. Biddle A, Stewart L, Blanchard J, Leschine S. 2013. Untangling the genetic basis of fibrolytic specialization by Lachnospiraceae and Ruminococcaceae in diverse gut communities. Diversity 5:627–640.
27. Stanley D, Geier MS, Denman SE, Haring VR, Crowley TM, Hughes RJ, Moore RJ. 2013. Identification of chicken intestinal microbiota correlated with the efficiency of energy extraction from feed. Vet Microbiol 164:85–92.
28. Torok VA, Hughes RJ, Mikkelsen LL, Perez-Maldonado R, Balding K, MacAlpine R, Percy NJ, Ophel-Keller K. 2011. Identification and characterization of potential performance-related gut microbiotas in broiler chickens across various feeding trials. Appl Environ Microbiol 77:5868–5878.
29. Yu Y, Lee C, Kim J, Hwang S. 2005. Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. Biotechnol Bioeng 89:670–679.
30. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541.

31. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. 2011. Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12:38.
32. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196.
33. Schloss PD, Handelsman J. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71:1501–1506.
34. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072.
35. Clarke KR. 2006. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18:117–143.

Appendix 2: List of publications not included in this thesis produced from work performed as a graduate student at the University of Wisconsin – Madison

A2.1 References

1. Omoniyi LA, Jewell KA, Isah OA, **Neumann AP**, Onwuka CF, Onagbesan OM, Suen G. **2014**. An analysis of the ruminal bacterial microbiota in West African Dwarf sheep fed grass- and tree-based diets. J Appl Microbiol 116:1094–105.
2. Mohammed R, Brink GE, Stevenson DM, **Neumann AP**, Beauchemin KA, Suen G, Weimer PJ. **2014**. Bacterial communities in the rumen of Holstein heifers differ when fed orchardgrass as pasture vs. hay. Front Microbiol 5:689.
3. Burnet MC, Dohnalkova AC, **Neumann AP**, Lipton MS, Smith RD, Suen G, Callister SJ. **2015**. Evaluating Models of Cellulose Degradation by *Fibrobacter succinogenes* S85. PLoS One 10(12):e0143809.