Characterization of the ruminal bacterial microbiota in relation to host parameters

by

Kelsea Anne Jewell

A dissertation submitted in partial fulfillment of

The requirements for the degree of

Doctor of Philosophy

(Microbiology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2014

Date of final oral examination: 7/15/14

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

Garret Suen, Assistant Professor, Bacteriology
Jeri Barak-Cunningham, Assistant Professor, Plant Pathology
Brian Fox, Professor, Biochemistry
Heidi Goodrich-Blair, Professor, Bacteriology
Paul Weimer, Associate Professor, Bacteriology

Abstract

The ruminants are a group of hindgut-fermenting animals of great agricultural importance, and include cattle and sheep. The ability of a ruminant to digest cellulose-rich feedstock is dependent on a rich ruminal microbial community, the complexity and necessity of which has long been recognized, but only partially understood. Ruminal bacteria, in particular, are responsible for the conversion of feedstock into a mixture of volatile fatty acids that are used by the host ruminant for milk production and body maintenance. Because of the dependence of the ruminant upon bacterial fermentation it is possible that the efficiency of feed conversion into milk products is strongly and directly impacted by the ruminal bacterial community structure. Further, modifications of the host diet can negatively or positively impact the membership and abundance of the ruminal bacterial community. In Chapters 2 and 3 of this thesis I present work investigating the correlation of the total ruminal bacterial community of lactating Holstein cows with host efficiency, and in both chapters I conclude that there is a strong and measurable distinction between the communities of higher and lower efficiency cows. In Chapter 2 I demonstrate this correlation with a group of eight multiparous, adult cows. In Chapter 3 I expand upon the work of Chapter 2 with a group of 14 cows sampled over the course of their first two full lactation cycles, with a repetition of the correlation in community structure and further results showing the impact of lactation stage on community diversity, richness, and membership. In Chapter 4 I use West African Dwarf sheep to demonstrate the impact of supplementing a basal grass diet of *Panicum maximum* with tree-based browse from *Albizia saman*, *Bridelia micrantha, Ficus sur,* or *Gmelina arborea* on the ruminal bacterial community.

Acknowledgements

First and foremost I thank my mentor Dr. Garret Suen for his trust and generosity, and Cathy Davis-Gray for her assistance and sympathy. Both gave me a hand up when the rest of the world was kicking me in the teeth, and I will never forget that. Everyone in the Suen lab has supported, critiqued, enthused, and commiserated in turns throughout my thesis work. In particular I thank my high school Youth Apprenticeship students Liz Strassman and Caroline McCormick for being enthusiastic and intelligent mentees who never complained when I made them help me in addition to completing their own research. And thank you, Jake Breaker, for volunteering to live at the farm over the summer and caring for our cows!

It would have been impossible for me to complete my two-year sampling study without the active support and encouragement of my husband, Dr. Rembrandt Haft, who never complained about the early alarms and often accompanied me on pre-dawn drives to the farm, and Doug and Aaron at Foreign Car Specialists for helping me buy and maintain Midge, the Cow Car. I thank Midge for getting me to the farm and back even when I couldn't see the highway for the snow, never failing to start in sub-zero weather, and having both an over-active heater and air-conditioner. I thank the herd of the USDA Dairy Forage Research Farm for teaching me how to work with sleepy cows in temperatures from -20 to 90F, and everyone at the Farm for their unflagging aid and goodwill, especially Kurt Pickar and Ron Skoyen. Of the cows in my studies, I wish to thank 4255 for being the ultimate sweetheart and 4288 for her many schemes and plots.

I thank my parents and brother for their active encouragement of me throughout my many years of school and research, and the Skullz of MDTP for the many Core Lunches, adventures, and unflagging support. Last, but never least, I thank my committee for their thoughts and support over the years.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE

Introduction

Many herbivorous mammals are ruminants, with the most conspicuous and agriculturally important members of this group being cattle and sheep. Ruminant growth and milk production rely upon fermentation products created by a microbial consortium present in the primary digestive organ known as the rumen. Thus the agricultural worth of a ruminant depends upon the ruminal microbiota, and much research has been performed to enumerate the diversity of ruminal microbes. Such descriptive microbiology has only recently been coupled to the technological advances necessary to begin finding correlations between ruminal community membership, fermentation products, and host performance metrics. The research presented in this dissertation reveals new information about the relationships between ruminal bacterial community, host agricultural performance, and diet in sheep and cattle. This research focuses on three main questions: Is there a stable, shared microbiota? What elements, if any, of the microbiota correlate with host production metrics? To what degree do major dietary changes impact the microbiota of the rumen?

The process of ruminant digestion

Bacteria in the guts of mammals convert complex polysaccharides, including extremely recalcitrant substrates such as cellulose, into simple sugars or other compounds for use in bacterial fermentation; released sugars and fermentation products are taken up through the host intestinal wall [1]. Despite this commonality there is a wide range of microbial community and fermentation strategies known to exist [2]. The two major modes of mammalian digestion differ in the location of fermentative microbes within the host: the hindgut, where fermentation takes place in the distal gut after feed passage through a proximal acid stomach, and the foregut, where fermentation takes place in a proximal near-neutral stomach with feed later passing through a

distal acid stomach [3]. In foregut fermenters (commonly referred to as ruminants) feed particles repeatedly pass between the mouth and rumen (the process of rumination), then through the reticulum and omasum ruminal chambers and finally into the true acid stomach or abomasum, and the intestines (Fig. 1A). Ruminants are a diverse group of herbivores ranging from the agriculturally familiar cows, sheep and goats to the undomesticated giraffe, moose, and elk. The focus of much research, including the work presented here, has been upon the commercially dominant species of cattle and sheep.

As grazing or browsing herbivores, cattle and sheep eat a diet primarily of leaves and young stems that, after being broken apart by repeated mastication and mixed liberally with saliva, is swallowed into the rumen. Such a diet is high in cellulose (β-1,4-linked glucose chains with a repeating cellobiose unit of two glucose molecules), hemicelluloses (the most abundant of which is xylan, consisting of side-chains with assorted modifications on a backbone of β-1,4-linked xylose chain), and pectin (α-1,4-linked D-galacturonic acid backbone with α-1,2-rhamnose and a multitude of other sugars and side-chains). Host enzymes cannot degrade these polysaccharides into digestible forms, so ruminants rely on microbial enzymes to create a mixture of host-available sugar monomers and oligomers, volatile fatty acids (VFAs, also referred to as short chain fatty acids), hydrogen, and carbon dioxide [4], the pool of which serves as metabolic precursors [5].

The entire ruminal microbial community includes bacteria, fungi, protists, and methanogenic archaea (Fig. 1B), but the microbes performing plant fiber digestion are anaerobic fungi in the phylum Neocallimastigomycota and a wide variety of bacteria [5]. Plant leaf and stem particles, once in the rumen, are colonized at breakages by both motile fungal zoospores [6]

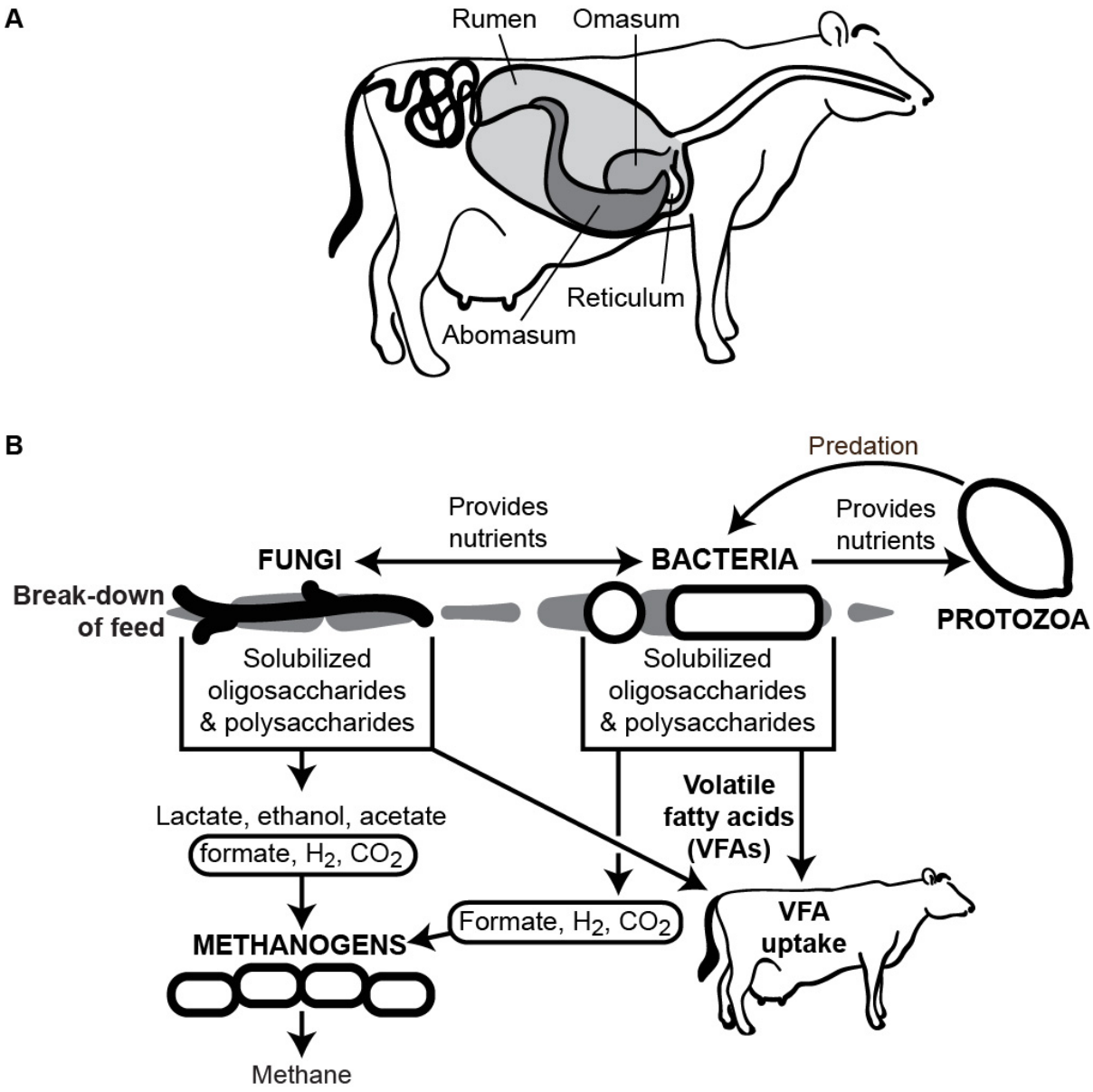

Figure 1. Simplified diagram of the ruminal microbiota and digestive system. (A) The ruminal system is composed of the reticulum, omasum, rumen, and abomasum. Fermentation primarily takes place in the rumen, while the abomasum is the acid or "true" stomach. (B) Both anaerobic fungi and bacteria digest cellulosic feed to produce solubilized oligosaccharides and polysaccharides that are used by other bacteria, volatile fatty acids used by the host, and fermentation by-products used by methanogens. Protozoa act as predators on ruminal bacteria. Adapted from [1,3,7].

and multiple fiber-adherent cellulolytic bacterial species [8]. After germination the fungal zoospores produce hyphae capable of further physical disruption of the plant cell walls in addition to enzymatic digestion, with further fiber digestion performed by bacteria [9]. Of the other ruminal microorganisms, protozoa are not cellulolytic and can negatively impact the activity of ruminal fungi and bacteria while archaeal methanogens act primarily as hydrogen sinks [7]. Methane, the most infamous byproduct of ruminal fermentation, is primarily expelled through eructation by the cow. Finally, there are a wide variety of non-cellulolytic bacteria in the rumen utilizing and transforming cellulose break-down products, which are discussed in more detail below.

The ruminal bacterial community is diverse and potentially host-driven

Microbial populations differ widely among the incoming feed, the rumen, the large intestines, and the small intestines, likely due to differences in environmental parameters (pH, water availability, oxygen tension, *etc.*) [10]. Despite the fluxes of feed, water, gas exchange, and substrate churning within the rumen [5], several bacterial phyla are consistently observed in association with ruminants: the Bacteroidetes (particularly the genus *Prevotella*), Firmicutes (especially the genera *Butyrivibrio*, *Megasphaera*, *Ruminococcus*, *Selenomonas*, *Streptococcus*), Proteobacteria (*Succinomonas*) and Fibrobacteres (*Fibrobacter*). Within these major groups, bacterial community variation between individual animals is high, even for cows on the same diet [11].

The high between-animal variation suggests that each ruminant establishes, as it develops, a ruminal community specific to itself. This impact of the host on the ruminal community may be responsible for the high degree of total community variation observed

between animals in identical conditions [12,13,14]. As a part of this, community structure is resistant to long-term perturbation. In a rumen-swapping experiment, where the contents of the rumen are removed, the rumen washed, and the contents replaced with those from another cow, it was seen that the total community structure swiftly shifted back to that originally present in that particular cow [12]. It is possible that this resistance to perturbation is due to host genetics, with the cow's specific physiology selecting, in combination with environmental factors, an animal-specific ruminal population in addition to a general population of bacteria to be found among all animals. In one three-year diet-shift study of steers, it was found that roughly half of the animals had their residual feed intake values (RFI, a measure of the difference in actual feed intake to the expected intake for body maintenance and growth, with lower values indicating increased efficiency) change depending on diet [15], suggesting that this fraction of animals were unable, either due to host or microbial factors, to adapt to fully digest both feeds. This idea is supported by work done by Hernandez-Sanabria *et al.*, where they found that steers with "middle" RFI values could be shifted by diet, but that steer with RFI-ranked extremes of high or low efficiency were resistant to dietary effects [16]. It is possible that for these animals there is either a strong selective pressure within the rumen for a particular microbial community, or that despite alterations in that community, the host animal is unable to take advantage of improved feed digestion.

The ruminal community is dynamic and is influenced by diet

The ruminal community can be expected to have a dynamic response to short-term alterations due to host behavior (such as drinking and eating) and long-term conditions (such as diet composition or host aging). Studies of ruminal microbiota dynamics have to date focused

upon diet composition [16,17,18,19]. Strong dietary impacts on the ruminal bacterial community are seen when there are changes in the dietary ratio of forage (dominated by cellulose and lignin [20]) to grains (dominated by starch [21]). Starch (α-1,4-linked glucose) is digested both by both ruminal and host amylases and the subsequent fermentation typically yields substantial amounts of lactate [22]. Thus increasing grain levels results in lowered populations of primary cellulose degrading bacteria such as *Fibrobacter succinogenes*, *Ruminococcus flavefaciens*, and *Butyrivibrio fibrisolvens* [18,23], and increased populations of lactate-utilizing bacteria, especially *Megasphaera elsdenii* [24-25]. Additionally, dietary supplementation with various polyunsaturated fatty acids (PUFA) increases the populations of lipolytic bacteria such as *Anaerovibrio lipolytica*, and further decreases *F. succinogenes* and *B. fibrosolvens* populations [24].

Ruminal fermentation products are required for milk production

From an agricultural standpoint the chief importance of a cow is her ability to produce a large volume of high-quality milk on minimal feed. High-quality milk is rich in proteins and fats, and its production is strongly affected by microbial metabolism. To emphasize the meaning behind "large volume," a single cow can produce nearly twice her bodyweight (for Holsteins this can be 750-820 kg) in milk fat alone over the course of a single lactation [25]. The standard lactation cycle is 305 days [26], with lactation beginning at the same time as calving. After the birth of her first calf, at approximately 24 months of age, the heifer is termed a cow, and on a modern farms will be kept for an average of 2.5 lactation cycles [27]. Early stage lactation (< 100 days into a lactation cycle) relies upon the mobilization of host tissues for energy through gluconeogenesis, while middle and late stages of lactation rely on VFAs as metabolic precursors

and replenishes the host tissues [28]. Thus it is important agriculturally, as well as scientifically, to better understand the host-microbe interactions that result in high-volume, high-quality milk production. Broadly, milk proteins come from digestion of feed and microbial proteins, while milk fats come from the major ruminal fermentation products.

The fermentation products that most influence milk production are the volatile fatty acids (VFAs) acetate, butyrate, propionate, and the compound lactate (for a recent review, see [29]); the production of these is dependent upon the ruminal microbes present and on the composition of the feed stock (Fig. 2A). Generally, cellulose-rich feeds are fermented primarily to acetate and butyrate, which promote production of high-quality milk, while starch-rich feeds are fermented to lesser amounts of acetate and butyrate and larger amounts of propionate and lactate; large amounts of propionate and lactate can reduce production of high-quality milk (Fig. 2B) [30,31]. Ruminal microorganisms also influence milk production through biohydrogenation of unsaturated fatty acids and the production of conjugated linoleic acid (CLA and 18:3 isomers) as intermediates and byproducts [31, 33-36]. CLAs play roles both in inhibiting milk fat synthesis and as secreted elements in milk fat [32,33,34] (Fig. 2B).

VFAs are absorbed through the ruminal wall at rates that dependent on ruminal concentration and host physiology [35]. One of the key VFAs is butyrate, which is converted to β-hydroxybutyrate as it passes from the rumen into the cow's tissues, and as β-hydroxybutyrate is becomes the precursor for approximately half of all milk odd- and branched-chain fatty acids (OBCFAs) of four carbons or longer [31]. Fatty acid chain length in cow milk is from 4-18 carbons, with the bulk of chains being 18:1, 16:1, 4:0, and 14:0 fatty acids [36]. Odd-chain fatty (OCFAs) acids are also synthesized from propionate and valerate, and branched-chain fatty acids (BCFAs) can also be synthesized from valine, leucine, isoleucine, and isobutyric, isovaleric, and

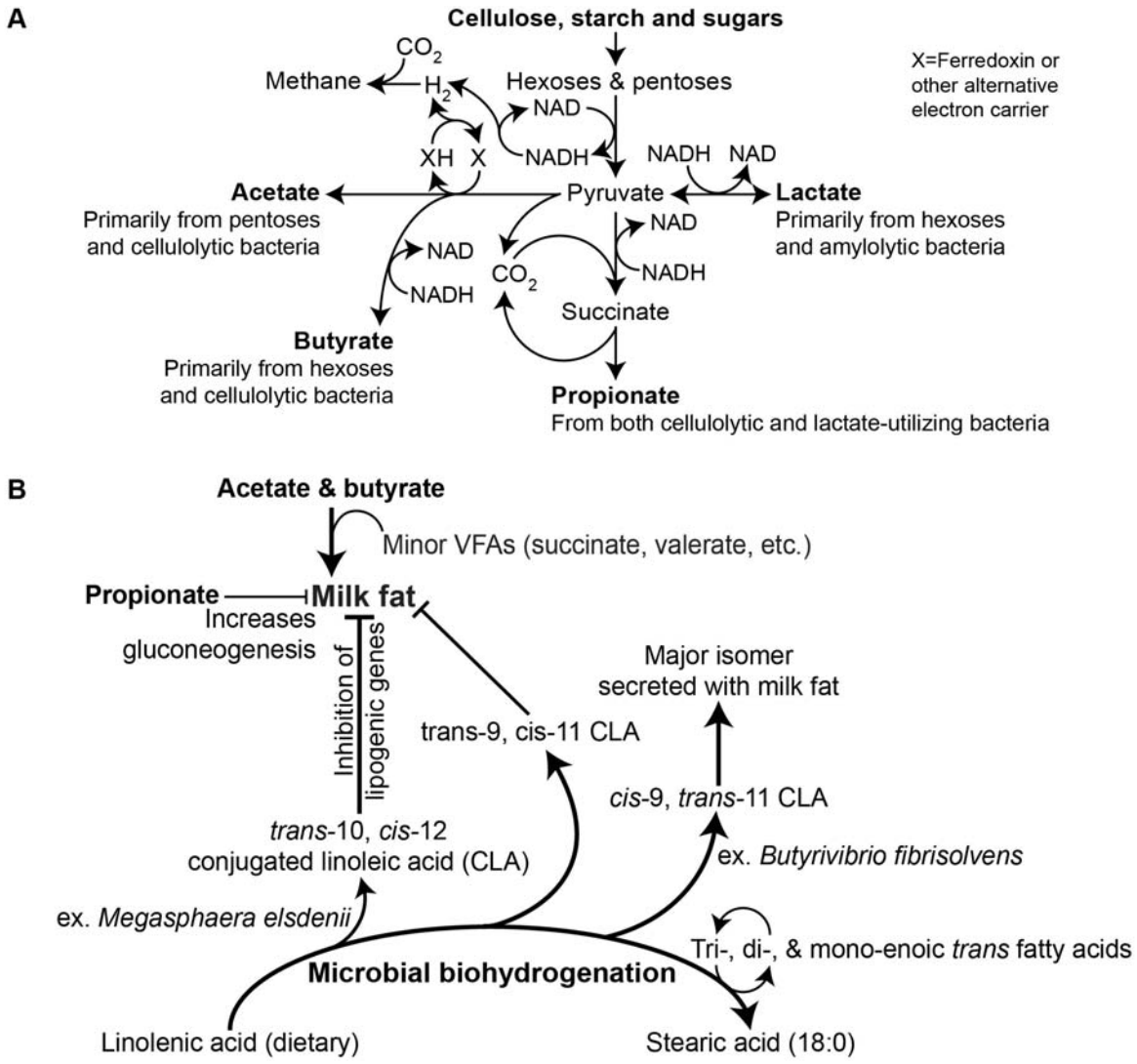

Figure 2. The ruminal system ferments feed into volatile fatty acids (VFAs) used in milk synthesis. (A) Cellulose, starch and sugars are fermented by microbes in the rumen into the principle VFAs acetate, butyrate, succinate, and propionate, and into methane and lactate. Adapted from [37]. (B) Simplified synthesis of milk fat as it is impacted by VFAs and bacterial biohydrogenation products, with line weight indicating relative degree of impact. Adapted from [32-33, 36-38].

2-methylbutyric acids [29]. Body fat can be used as a precursor for OBCFAs, but accounts for less then 10% of milk fatty acids, except in cases of massive energy deficits [36].

The profile of VFAs available to the ruminant for OBCFA synthesis, and the concentration of inhibitory CLAs, depends on the ruminal community and each animal's uptake and processing rate. Direct infusions of butyrate in the rumen can increase milk fat [38] and protein production [38,39], suggesting that increasing microbial output of desirable VFAs could have a positive impact on milk component production until host uptake is saturated or metabolism is unbalanced [38],. Increasing dietary starch encourages the growth of amylolytic over cellulolytic bacteria, thereby shifting the major fermentation products of acetate and butyrate to a mixture of propionate, butyrate, acetate, and lactate; this shift in VFAs results in reductions in the concentrations of some milk BCFAs and an increase in inhibitory CLAs [40,41,42].

Protein and lactose are also impacted by the presence and activity of ruminal microbes. Both fungi and bacteria within the rumen perform biohydrogenation of polyunsaturated fatty acids to create a variety of products used for milk protein and lactose synthesis and as a major energy source for body maintenance [29,33,43,44]. Additional milk proteins come from amino acids derived directly from the feed [45] or from proteolysis of microorganisms in the lower digestive tract [5]. Biohydrogenation also creates *trans*-10 *cis*-12 conjugated linoleic acid (*trans*-10 *cis*-12 CLA) [41,43], which along with *trans*-9 *cis*-11 CLA has been strongly connected with milk fat depression by decreasing acetate incorporation into fatty acids, depressing mRNA levels for lipid synthesis, and increasing adipose tissue (as opposed to mammary tissue) lipid synthesis [32,33,34].

Negative impacts of ruminal fermentation products

The positive impacts of the ruminal community have already been discussed in terms of breaking down complex carbohydrates into monomers and the production of fermentation products. In addition to these critically useful roles the ruminal community can cause host morbidity and mortality through the build-up of undesirable fermentation products via overgrowth of certain bacterial species. In dairy cows one of the most commonly problematic conditions is a reduction in milk fat production termed clinical milk fat depression (MFD); this condition can be induced by a combination of dietary starch, lactate, and propionate imbalances. Grains, which are high in starch [21], are fermented to lactate, which lowers the ruminal pH. Lowered pH can potentially lead to ruminal acidosis, the inhibition of microbial protein synthesis, and in extreme cases death [5]. In addition to these direct effects, there is an increase in the growth of the lactate-utilizing bacterium *Megasphaera elsdenii*, a known propionate producer [46] associated with milk fat depression [47,48] and ketosis [49]. High concentrations of ruminal propionate are associated with lowered milk fat levels and MFD [38,39,50,51], probably through the induction of gluconeogenesis and the resulting insulin-related decrease in blood levels of milk fat precursors [31]. Secondary morbid conditions triggered by high concentrations of dietary starch, and the resulting increase in *M. elsdenii*, include subacute and acute ruminal acidosis [47].

Despite being linked to these morbid conditions *M. elsdenii*, and the similarly lactate-utilizing *Selenomonas ruminantium*, has been investigated as a potential probiotic treatment for lactating dairy cows [50] and to aid in the adaptation of steer to high-starch diets [52]. It is currently unclear to what extent *M. elsdenii* is responding or actively contributing to ruminal conditions harmful to a cow's health and lactation. For *M. elsdenii*, and probably all ruminal

bacteria, the answer may lie not in the individual bacterium, but in its relation to the whole ruminal microbiota and host health.

The use of sequencing technology to study the ruminal microbiota

The interactions of ruminal microbes with their host, either negative or positive, may not lie in any one specific species. Instead, the pattern and membership of the entire ruminal community may be the driving force, with secondary impacts arising from nutrition, health, age, or other host variables. Therefore, I purposed to investigate the whole ruminal microbiota in dairy cows over multiple lactation cycles in order to ask questions regarding the total temporal shifts in the ruminal bacterial community as functions of time and host production metrics.

In order to examine the whole membership of the ruminal bacterial community I chose to use the culture-independent sequencing method of 454 pyrosequencing combined with the monitoring and analysis of host physiological parameters including feed consumption,body weight, ruminal chemistry, milk production, and milk composition. Our current knowledge of the ruminal community is based in large part on extensive culture work, community fingerprinting such as DGGE, and clone libraries. The recent rise in the usage of sequence-based investigations, including the 454 pyrosequencing platform, has underscored the complexity of the ruminal microbial microbiota by expanding upon the detection sensitivity of earlier methods. In pyrosequencing mixed templates of extracted DNA from the rumen are used for randomized amplification of chosen variable regions in the 16S rRNA gene; these amplicons, with sample origin identified by a PCR-added barcode, are pooled and sequenced in a single run. The resulting sequences, often 150,000+ per run and 150-700 bp in length depending on the selected

16S rRNA gene region, are compared to existing databases in order to obtain comparative measures of bacterial diversity and identity.

As with all methods, there are technical difficulties with relying upon sequencing from complex, environmentally-derived DNA templates and from ruminal contents in particular. First, although sequence-based results have supported and expanded on the previously observed dominance of the Bacteroidetes, Firmicutes, Proteobacteria, and in particular the genus *Prevotella* [17,53,54,55,56,57,58,59] they have often failed to detect the Fibrobacteres [53,54,56,57,58], probably due to known difficulties in amplifying even pure culture-derived *F. succinogenes* DNA [23]. There is also the loss of some bacterial species due to the physical disruption steps required for separating fiber-adherent bacteria from rumen solids [60]. Finally, the creation of DNA amplicon libraries for sequencing will always be biased in that the GC content, and primer mismatches even for "universal" degenerate primers, will favor some 16S rRNA gene sequences over others [61]. Within these limitations, pyrosequencing and other large-scale sequencing methods are currently the best technology for detecting and analyzing ruminal bacterial community membership at a previously impossible level of detail and coverage.

The increasing number of sequences generated from the rumen have highlighted the variety and diversity of bacteria previously unknown to ruminal microbiology; many of these sequences remain currently unidentifiable to any named members of the Bacteria. By comparing the diversity of the conserved 16S rRNA gene in bacteria, it has been estimated that the rumen contains 300-400 distinct species of which perhaps 200 have been cultured [53]. Both numbers may yet grow as sequencing coverage and depth increase with improved technology, and as culture-based efforts are targeted towards these "unclassified" bacteria. Therefore, the function

of these sequence-identified bacteria can currently only be conjectured, but the ultimate impact and importance of the ruminal consortium on host performance has long been studied. The use of large-scale sequencing and other recent technologies allow for the creation of an increasingly fine-grained identification of the microbial taxa of greatest import in modifying that host performance.

Relating the ruminal bacterial community to host agricultural performance and diet

Given that the ruminal community is responsible for releasing or creating the compounds required for animal growth, maintenance, and milk production, it is reasonable to look for links between the community composition and measurable host factors. Most studies investigating links between the ruminal community and host efficiency have used growth-related metrics, especially in beef cattle [16,62,63]. For dairy animals, the chief measurable characteristic is efficiency of milk production. Dairy production efficiency as linked to the ruminal community using low-resolution techniques such as amplicon length heterogeneity PCR [64], or focusing on a specific condition such as ruminal acidosis [65]. Very recent work by Jami *et al.* used lactating cows and found that there was a link between the relative proportion of Bacteroidetes to Firmicutes and milk fat percentage [66]. This study did not find any significant correlation between the entire bacterial community and production efficiency as measured using a technique known as residual feed intake (RFI) [66]. Because the animals in this study did not display great ranges in production metrics (for example, milk fat production was 1.37 ± 0.04 kg/d across all 15 cows [66]), it is possible that there was not enough variation between animals to establish such a correlation. In addition, the rumen contents were only sampled for a single day, making it impossible to tell if there were long-term trends linking community and performance. Thus, the

impact of the ruminal bacterial community on host efficiency, especially over time, was an open question that the work of this thesis was designed to answer.

To answer this question I used two systems, the Holstein dairy cow and the West African Dwarf meat sheep. I formed a major hypotheses, that the first that the ruminal bacterial community is different between high and low production animals, and a minor hypothesis, that there is a wide latitude of total community composition associated with "healthy" animals. I have addressed these hypothesis in my dissertation with three main questions: (1) What are the shared, common ruminal bacteria? (2) To what degree is milk production efficiency correlated to members of the ruminal bacterial community? (3) How is the ruminal bacterial community impacted by major dietary shifts? For all three questions I used the same technical approach of collecting ruminal liquids and solids, extracting total DNA from these samples, and identifying the bacteria present by 454 pyrosequencing of the 16S rRNA gene variable regions V6-V8. By using the same sequencing technique and gene region, I increased my ability to compare data across multiple studies. To support my sequencing data I also, where possible, sampled the ruminal fluids for VFA concentration, measured feed consumption and dietary sorting, tracked host health, and quantified milk production and composition.

To address my first two questions I needed a way to measure efficiency, but there is no clear consensus yet established for efficiency reporting in dairy cattle, although many have been proposed [67]. I chose to use two methods for measuring and reporting efficiency, basing my choice primarily on applicability and feasibility within each trial. For my first study (Chapter 2), I followed a group of cows of mixed-parity and lactation stage for a relatively short period (less than one month). Thus the efficiency calculations included multiple energetic modifying terms [28] for pregnancy, mastitis [68], bodyweight with tissue mobilization and body maintenance,

milk composition, and feed consumption and selective feed refusal. The final milk production efficiency was a relative measure, as it was modified by the many major physiological differences between the cows.

For my second study (Chapter 3) I used a simpler method for measuring milk production efficiency. As I was following the same group of heifers from the same birth cohort, and because I was only comparing values collected on the same days in milk for each animal, I did not need to include terms to make adjustments for pregnancy stage. Because I did not have weight data for the heifers during their first lactation cycle I could not include bodyweight, tissue mobilization, or body maintenance terms. For these reasons I chose to use gross feed efficiency (GFE) [67], which is energy corrected milk divided by dry matter intake (ECM/DMI). ECM is calculated from the milk composition (fat, protein, and lactose) and yield (mass of milk) [22] and DMI is calculated from the dry mass of feed consumed. Using GFE and relative feed efficiency I could begin to answer my questions regarding correlations and changes within the ruminal bacterial community as they relate to milk production.

I addressed my first question, that of the shared ruminal bacterial community, using both Holstein dairy cows and heifers, and West African Dwarf (WAD) meat ewes (Chapters 2-4). The group of cows were of mixed parity and life history, the heifers were of the same parity and life history, and the ewes were undergoing multiple feeding trials. For my second question, where I was particularly interested in correlating milk production efficiency with changes in the ruminal microbiota, I used the same two groups of Holsteins (Chapters 2-3). By using two different groups I was able to ask my questions in the context of a mixed-parity group at a single point in time (Chapter 2) as well as with a single-parity group over the course of their first two full lactation cycles (Chapter 3). With the long-term group I included time effects relating to

early, middle, and late points in their lactation cycles in order to examine the impact of the natural shift in a cow during the course of each lactation from mobilization of host tissues (early lactation) to ruminal VFAs (mid-to-late lactation) as energy sources. In both chapters I found that there was a measurable and repeated correlation of the bacterial ruminal community membership with measures of production efficiency. Further, in Chapter 2 I show that the diversity and richness of the community changed over time, most strikingly with a steady increase in the diversity of the ruminal liquids as the animals matured.

For the third question I tracked the weight gain and ruminal community change in WAD sheep during a feeding trial that tested the feasibility of using tree-based instead of grass-based feed (Chapter 4). In that study I found that, despite all of the sheep beginning with highly idiosyncratic ruminal bacterial communities, there were strong patterns of change in those communities based on the type of feed substitution. All parts of my work were designed to increase our knowledge of the ruminal bacterial community's role in shaping, and being shaped by, the host animal.

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. Nature Reviews Microbiology 6: 121-131.
2. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, et al. (2008) Evolution of mammals and their gut microbes. Science 320: 1647-1651.
3. Dehority BA (2003) Rumen Microbiology. Thrumpton, Nottingham: Nottingham University Press.
4. Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, et al. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331: 463-467.
5. Russell JB (2002) Rumen microbiology and its role in ruminant nutrition. Ithaca, NY: Cornell University. 120 p.
6. Edwards JE, Kingston-Smith AH, Jimenez HR, Huws SA, Skøt KP, et al. (2008) Dynamics of initial colonization of nonconserved perennial ryegrass by anaerobic fungi in the bovine rumen. FEMS Microbiology Ecology 66: 537-545.
7. Cheng YF, Edwards JE, Allison GG, Zhu W-Y, Theodorou MK (2009) Diversity and activity of enriched ruminal cultures of anaerobic fungi and methanogens grown together on lignocellulose in consecutive batch culture. Bioresource Technology 100: 4821-4828.
8. Edwards JE, Huws SA, Kim EJ, Kingston-Smith AH (2007) Characterization of the dynamics of initial bacterial colonization of nonconserved forage in the bovine rumen. FEMS Microbiology Ecology 62: 323-335.
9. Lee SS, Ha JK, Cheng KJ (2000) Relative contributions of bacteria, protozoa, and fungi to *in vitro* degradation of Orchard Grass cell walls and their interactions. Applied and Environmental Microbiology 66: 3807-3813.
10. de Oliveira MNV, Jewell KA, Freitas FS, Benjamin LA, Tótola MR, et al. (2013) Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. Veterinary Microbiology 164: 307-314.
11. Welkie DG, Stevenson DM, Weimer PJ (2010) ARISA analysis of ruminal bacterial community dynamics in lactating dairy cows during the feeding cycle. Anaerobe 16: 94-100.
12. Weimer PJ, Stevenson DM, Mantovani HC, Man SLC (2010) Host specificity of the ruminal bacterial community in the dairy cow following near-total exchange of ruminal contents1. Journal of Dairy Science 93: 5902-5912.
13. Li M, Penner GB, Hernandez-Sanabria E, Oba M, Guan LL (2009) Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen. Journal of Applied Microbiology 107: 1924-1934.
14. Omoniyi LA, Jewell KA, Isah OA, Neumann AP, Onwuka CFI, et al. (2014) An analysis of the ruminal bacterial microbiota in West African Dwarf sheep fed grass and tree-based diets. Journal of Applied Microbiology 116: 1094-1105.
15. Durunna ON, Mujibi FDN, Goonewardene L, Okine EK, Basarab JA, et al. (2011) Feed efficiency differences and reranking in beef steers fed grower and finisher diets. Journal of Animal Science 89: 158-167.
16. Hernandez-Sanabria E, Goonewardene LA, Wang Z, Durunna ON, Moore SS, et al. (2012) Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. Applied and Environmental Microbiology 78: 1203-1214.

17. de Menezes AB, Lewis E, O'Donovan M, O'Neill BF, Clipson N, et al. (2011) Microbiome analysis of dairy cows fed pasture or total mixed ration diets. FEMS Microbiology Ecology 78: 256-265.
18. Fernando S, Purvis II H, Najar F, Sukiharnikov L, Krehbiel C, et al. (2010) Rumen microbial population dynamics during adaptation to a high-grain diet. Applied and Environmental Microbiology 76: 7482-7490.
19. Stiverson J, Morrison M, Yu Z (2011) Populations of select cultured and uncultured bacteria in the rumen of sheep and the effects of diets and ruminal fractions. International Journal of Microbiology 2011: doi:10.1155/2011/750613.
20. Broderick GA (1985) Alfalfa silage or hay versus corn silage as the sole forage for lactating dairy cows. Journal of Dairy Science 68: 3262-3271.
21. Herrera-Saldana R, Huber J, Poore M (1990) Dry matter, crude protein, and starch degradability of five cereal grains. Journal of Dairy Science 73: 2386-2393.
22. Huntington GB (1997) Starch utilization by ruminants: from basics to the bunk. Journal of Animal Science 75: 852-867.
23. Tajima K, Aminov R, Nagamine T, Matsui H, Nakamura M, et al. (2001) Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR. Applied and Environmental Microbiology 67: 2766-2774.
24. Liu S, Bu D, Wang J, Liu L, Liang S, et al. (2011) Effect of incremental levels of fish oil supplementation on specific bacterial populations in bovine ruminal fluid. Journal of Animal Physiology and Animal Nutrition DOI: 10.1111/j.1439-0396.2010.01113.x.
25. Bauman DE, Harvatine KJ, Lock AL (2011) Nutrigenomics, Rumen-Derived Bioactive Fatty Acids, and the Regulation of Milk Fat Synthesis. Annual Review of Nutrition 31: 299-319.
26. Dekkers J, Ten Hag J, Weersink A (1998) Economic aspects of persistency of lactation in dairy cattle. Livestock Production Science 53: 237-252.
27. EPA (2012) Lifecycle Production Phases
28. National Research Council (2001) Nutrient requirements of dairy cattle: National Academy Press, Washington DC.
29. Fievez V, Colman E, Castro-Montoya JM, Stefanov I, Vlaeminck B (2012) Milk odd- and branched-chain fatty acids as biomarkers of rumen function—An update. Animal Feed Science and Technology 172: 51-65.
30. Baldwin VI RL, Allison M (1983) Rumen metabolism. Journal of Animal Science 57: 461-477.
31. Bauman D, Griinari J (2003) Nutritional regulation of milk fat synthesis. Annual Review of Nutrition 23: 203-227.
32. Bernard L, Leroux C, Chilliard Y (2008) Expression and nutritional regulation of lipogenic genes in the ruminant lactating mammary gland. In: Bösze Z, editor. Bioactive Components of Milk: Springer New York. pp. 67-108.
33. Lee Y-J, Jenkins TC (2011) Biohydrogenation of linolenic acid to stearic acid by the rumen microbial population yields multiple intermediate conjugated diene isomers. The Journal of Nutrition 141: 1445-1450.
34. Perfield II JW, Lock AL, Griinari JM, Sæbø A, Delmonte P, et al. (2007) *Trans*-9, *cis*-11 conjugated linoleic acid reduces milk fat synthesis in lactating dairy cows. Journal of Dairy Science 90: 2211-2218.

35. Dijkstra J, Boer H, Van Bruchem J, Bruining M, Tamminga S (1993) Absorption of volatile fatty acids from the rumen of lactating dairy cows as influenced by volatile fatty acid concentration, pH and rumen liquid volume. British Journal of Nutrition 69: 385-396.
36. Bauman DE, Griinari JM (2001) Regulation and nutritional manipulation of milk fat: low-fat milk syndrome. Livestock Production Science 70: 15-29.
37. Russell JB, Rychlik J (2001) Factors that alter rumen microbial ecology. Science 292: 1119-1122.
38. Miettinen H, Huhtanen P (1996) Effects of the ratio of ruminal propionate to butyrate on milk yield and blood metabolites in dairy cows. Journal of Dairy Science 79: 851-861.
39. Huhtanen P, Miettinen H, Ylinen M (1993) Effect of increasing ruminal butyrate on milk yield and blood constituents in dairy cows fed a grass silage-based diet. Journal of Dairy Science 76: 114-1124.
40. Vlaeminck B, Fievez V, Tamminga S, Dewhurst RJ, van Vuuren A, et al. (2006) Milk odd and branched-chain fatty acids in relation to the rumen fermentation pattern. Journal of Dairy Science 89: 3954-3964.
41. Nielsen T, Sejrsen K, Andersen H, Lund P, Straarup E (2004) Effect of silage type and energy concentration on conjugated linoleic acid (CLA) in milk fat from dairy cows. Journal of Animal and Feed Science 13: 697-700.
42. Shingfield K, Reynolds C, Lupoli B, Toivonen V, Yurawecz M, et al. (2005) Effect of forage type and proportion of concentrate in the diet on milk fatty acid composition in cows given sunflower and fish oil. Animal Science 80: 225-238.
43. Nam I, Garnsworthy P (2007) Biohydrogenation of linoleic acid by rumen fungi compared with rumen bacteria. Journal of Applied Microbiology 103: 551-556.
44. van Houtert MFJ (1993) The production and metabolism of volatile fatty acids by ruminants fed roughages: A review. Animal Feed Science and Technology 43: 189-225.
45. Wright TC, Moscardini S, Luimes PH, Susmel P, McBride BW (1998) Effects of rumen-undegradable protein and feed intake on nitrogen balance and milk protein production in dairy cows. Journal of Dairy Science 81: 784-793.
46. Counotte GHM, Prins RA, Janssen RHAM, deBie MJA (1981) Role of *Megasphaera elsdenii* in the fermentation of dl-[2-^{13}C]lactate in the rumen of dairy cattle. Applied and Environmental Microbiology 42: 649-655.
47. Khafipour E, Li S, Plaizier JC, Krause DO (2009) Rumen microbiome composition determined using two nutritional models of subacute ruminal acidosis. Applied and Environmental Microbiology 75: 7115-7124.
48. Palmonari A, Stevenson DM, Mertens DR, Cruywagen C, Weimer PJ (2010) pH dynamics and bacterial community composition in the rumen of lactating dairy cows. Journal of Dairy Science 93: 279-287.
49. Wang X, Li X, Zhao C, Hu P, Chen H, et al. (2012) The correlation between composition of the bacterial community and concentration of volatile fatty acids in the rumen during the transition period and ketosis in dairy cows. Applied and Environmental Microbiology.
50. Aikman P, Henning P, Humphries D, Horn C (2011) Rumen pH and fermentation characteristics in dairy cows supplemented with *Megasphaera elsdenii* NCIMB 41125 in early lactation. Journal of Dairy Science 94: 2840-2849.
51. Rook J, Blach C, Johnson V (1965) Further observations on the effects of intraruminal infusions of volatile fatty acids and of lactic acid on the yield and composition of the milk of the cow. British Journal of Nutrition 19: 93-99.

52. Drouillard J, Henning P, Meissner H, Leeuw K-J (2012) *Megasphaera elsdenii* on the performance of steers adapting to a high-concentrate diet, using three or five transition diets. South African Journal of Animal Science 42: 195-199.
53. Kong Y, Teather R, Forster R (2010) Composition, spatial distribution, and diversity of the bacterial communities in the rumen of cows fed different forages. FEMS Microbiology Ecology 74: 612-622.
54. Kim M, Morrison M, Yu Z (2011) Status of the phylogenetic diversity census of ruminal microbiomes. FEMS Microbiology Ecology 76: 49-63.
55. Thoetkiattikul H, Mhuantong W, Laothanachareon T, Tangphatsornruang S, Pattarajinda V, et al. (2013) Comparative Analysis of Microbial Profiles in Cow Rumen Fed with Different Dietary Fiber by Tagged 16S rRNA Gene Pyrosequencing. Current Microbiology 67: 130-137.
56. Stevenson DM, Weimer PJ (2007) Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Applied Microbiology and Biotechnology 75: 165-174.
57. Jami E, Mizrahi I (2012) Composition and similarity of bovine rumen microbiota across individual animals. PLOS ONE 7: e33306.
58. Callaway TR, Dowd SE, Edrington TS, Anderson RC, Krueger N, et al. (2010) Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing. Journal of Animal Science 88: 3977-3983.
59. Kittelmann S, Seedorf H, Walters W, Clemente JC, Knight R, et al. (2013) Simultaneous amplicon sequencing to explore co-occurence patterns of bacterial, archaeal, and eukaryotic microorganisms in rumen microbial communities. PLOS ONE 8: e47879.
60. Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, et al. (2013) Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. PLOS ONE 8: e74787.
61. Quail MA, Otto TD, Gu Y, Harris SR, Skelly TF, et al. (2012) Optimal enzymes for amplifying sequencing libraries. Nat Meth 9: 10-11.
62. Guan LL, Nkrumah JD, Basarab JA, Moore SS (2008) Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. FEMS Microbiology Letters 288: 85-91.
63. Carberry CA, Kenny DA, Han S, McCabe MS, Waters SM (2012) Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. Applied and Environmental Microbiology 78: 4949-4958.
64. Lettat A, Benchaar C (2013) Diet-induced alterations in total and metabolically active microbes within the rumen of dairy cows. PLOS ONE 8: e60978.
65. Mao SY, Zhang RY, Wang DS, Zhu WY (2013) Impact of subacute ruminal acidosis (SARA) adaptation on rumen microbiota in dairy cattle using pyrosequencing. Anaerobe 24: 12-19.
66. Jami E, White BA, Mizrahi I (2014) Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. PLOS ONE 9: e85423.
67. Connor EE, Hutchison JL, Olson KM, Norman HD (2012) Triennial Lactation Symposium: Opportunities for improving milk production efficiency in dairy cattle. Journal of Animal Science 90: 1687-1694.

68. DRMS (2013) The DHI Glossary, August 2013. Raleigh, NC: Dairy Records Management Systems. pp. 20.

CHAPTER TWO

Ruminal bacterial community composition correlates in lactating dairy cattle with feed efficiency

Kelsea A. Jewell, Christine L. Odt, David M. Stevenson, Paul J. Weimer, and Garret Suen

"Ruminal bacterial community composition correlates in lactating dairy cattle is divergent by feed efficiency"

A version of this work has been submitted for publication to *Applied Microbiology and Biotechnology*, June, 2014

Animal samples were collected by KJ, CO, DS, PW, and GS. KJ processed the rumen samples, performed the sequencing and sequence analyses, and wrote the manuscript. CO and DS processed the feed samples. CO performed HPLC analysis of the VFA samples. PW performed additional statistical analyses. PW and GS designed the experiments and assisted with writing the manuscript.

Abstract

Milk production in cattle is dependent on ruminal bacteria that produce volatile fatty acids (VFAs) that are converted by cows into milk components. We hypothesized that specific ruminal bacterial communities would be conserved and associated with high-efficiency (HE) or low-efficiency (LE) animals and rumen fluid chemistry (VFAs and succinate). We tested this hypothesis by measuring the total ruminal bacterial community composition, VFA profiles, and feed efficiency in eight fistulated Holstein dairy cows. Ruminal solids and liquids were analyzed for community composition by 454 pyrosequencing of the V6-V8 region of the 16S rRNA gene. Feed efficiency was calculated using milk production and composition, feed consumption, stage of lactation, and pregnancy metrics. The total community composition differentiated between liquid and solid phases, and between animals in correlation with relative efficiency classification for six of the eight animals. We identified shared sets of operational taxonomic units (95 % OTUs) within efficiency groups and among all cows. We also identified specific genera and individual OTUs responsible for discriminating between HE (*Butyrivibrio*, *Moryella*, *Paludibacter*, and *Prevotella*) and LE animals (*Lachnospira*, *Prevotella*, and *Sharpea*). Propionate concentrations were also significantly higher in LE cows ($P = 0.033$). We conclude that ruminal bacterial communities are associated with feed efficiency in dairy cows, and that these differences cannot be predicted by ruminal VFA profiles.

Introduction

Dairy cattle are one of the most economically important agricultural animals worldwide. For example, in the USA, annual cow milk production is valued at 50 billion dollars [3,4]. There is a growing understanding that milk production is influenced by a number of factors including diet, environmental conditions, animal genetics, and associated ruminal microbial communities [5]. The specialized ruminal microbiota is responsible for the conversion of plant matter into volatile fatty acids (VFAs) that are used by the cow as nutrient sources for tissue growth, body maintenance, and milk synthesis. During lactation, acetate and butyrate are used as metabolic precursors for the production of milk fat, while propionate is used primarily for gluconeogenesis to satisfy energy demands for growth, lactation, and fetal development. In high concentrations, propionate is known to depress milk fat production [5,6]. The ruminal VFA profile is dependent upon a complex microbial community, with amylolytic bacteria producing propionate, butyrate, acetate, and lactate, while fibrolytic bacteria generate primarily acetate and butyrate [6]. Because VFAs are microbial products, decreased VFA levels in beef steers have been linked to shifts in microbial communities that result in poor feed conversion efficiency [7]. Therefore, an improved understanding of milk production and composition must consider the impact of the ruminal microbial community.

The use of fistulation in ruminants, which allow for direct access to the ruminal contents through a surgically implanted portal, combined with quantifiable host metrics and multiple methods for community profiling [8] presents a highly manipulable model system for investigating host-microbe interactions. Using this system, studies have shown that diet has a direct impact on the ruminal microbial community [9,10,11,12,13]. Moreover, the ruminal microbial community is known to impact host health [14], ruminal fermentation [12,15], and

feed conversion efficiency [7,13]. Recent studies using next-generation sequencing have partially elucidated the diversity and composition of the ruminal microbial community [16,17,18] at a level of coverage hitherto impossible. This technique is just beginning to be used in combination with individual production-related metrics and feed conversion efficiency [16] to clarify to what extent the ruminal bacterial community influences host performance or *vice versa*. To date there has been no published work investigating the potential correlation of the ruminal bacterial community with total milk production efficiency in dairy cows over multiple periods within a lactation, or over multiple lactation cycles.

In order to better understand the relationship of the ruminal community to host function, we compared ruminal bacterial populations from cows grouped by relative feed efficiency metrics. We hypothesized that cows with similar milk production efficiencies will have similar ruminal bacterial communities. We used ruminal liquids and solids for bacterial 16S rRNA pyrosequencing and VFA analysis. Importantly, our analysis involved both blind (without efficiency classification as a variable) and grouped (with animals assigned to high or low efficiency groups) methods. These data were used to establish correlations between bacterial community members and host function.

Materials and methods

Animal care. A group of 13 fistulated, lactating Holstein dairy cows (*Bos taurus*) were selected from the US Dairy Forage Research Center (USDFRC) farm herd (Prairie du Sac, WI) and used according to protocol A01104, as approved by the University of Wisconsin-Madison's Institutional Animal Care and Use Committee. All animals were fed once daily post morning milking with a total mixed ration (TMR: major components were alfalfa haylage, corn silage,

and high moisture shell corn, see Supplementary Table S1) [17] for at least two weeks prior to, and then during, sample collections. Feed refusal and TMR collections started one day prior to milk collections as morning milk production is dependent on feed consumed the previous day. No cows became ill during the study and ruminal pH and milk production values (volume and composition) for all cows were within expected normal ranges [18] (Supplementary Table S2). All animals had *ad libitum* access to water and were kept in tie-stalls in a single barn with other members of the USDFRC herd. Feed intake, refusal weights, and feed samples were collected daily for each cow over nine days following established methods [19]; milk samples were also collected for nine consecutive days during morning and evening milking and stored at 4 °C. Milk samples were submitted to AgSource Cooperative Services (Verona, WI) for near-infrared spectroscopic prediction [20] of percent milk fat, lactose, protein, non-fat solids, milk urea nitrogen, and somatic cell counts.

Pairwise selection of cows for feed efficiency. Because of the competing demands of lactation, pregnancy, growth, maintenance, and immunological function, a cow exhibits marked changes in feed efficiency during a lactation cycle and over her lifetime; thus it is difficult to compare efficiencies among cows that differ in age, size, and physiological state [21]. Consequently we elected to identify pairs of cows that displayed differences in overall feed efficiency at similar physiological states. From our group of 13 ruminally-fistulated cows, 4 pairs were selected that displayed within-pair similarities in parity (lactation number), pregnancy (days carrying calf, DCC), stage of lactation (days in milk, DIM), and within-pair numeric differences in feed efficiency. The other five cows did not display distinctive feed efficiencies when paired for parity, pregnancy, and stage of lactation. Feed efficiency was calculated to include five major and quantifiable contributions to the cow's energetics (all calculated in MJ): energy-corrected

Table S1. Feed ingredients and composition for total mixed ration (TMR) diet.

Ingredient	**% dry matter**
Alfalfa haylage	32.3
Corn silage	30.6
High-moisture shell corn	16.2
Dry corn	3.8
Roasted soybeans	8.0
Dried distillers grains	3.7
Canola meal	1.5
Soybean meal (48% crude protein)	1.5
Mineral and Vitamin mix	2.4
Composition	
Neutral detergent fiber	32.8
Nonfiber carbohydrate	28.6
Crude protein	16.0
Fat (ether extract)	4.4
Ash	5.7

milk (ECM) production; body weight change (ΔBW); maintenance (a function of body weight); gestation; and the energetic load of mastitis infection (the major immunological stress in high-producing dairy cows). Nonproductive energetic losses in manure and methane emissions were excluded from the calculations. These formulas are presented in Table 1 and are taken from [1,2]. ECM from the two daily milkings was summed to provide a separate daily production of ECM for each of the 9 d of milk sample collection. Cows were weighed on two successive days at the beginning and end of the milk sampling period, and body weight gain calculated as the difference between the mean body weights determined on the first two and last two days of the milk sampling period. The energetic cost of mastitis was determined from the conversion table for potential milk production loss based on somatic cell counts (SCC) in milk, determined at each milking [2]; these losses were averaged over the same 9 d milk sampling period. The aggregated energy demands of the five energetic components for each sampling day were divided by the dry matter intake (DMI, calculated as kg DM fed minus kg DM of refusals from the previous day) to calculate adjusted feed efficiency and expressed as MJ/kg DMI.

Ruminal sample collection and processing. All samples were collected in November of 2011. Ruminal solids and liquids were collected through the rumen fistula once daily prior to feeding on three successive days at the end of the 9-day milk sampling period. Ruminal liquids were strained through four layers of cheesecloth, and ruminal solids were squeezed thoroughly to remove liquid. All samples were transported on wet ice and frozen at -80 °C prior to DNA extraction, with 1.5 mL aliquots removed from the ruminal liquids and stored at -80 °C for volatile fatty acid (VFA) analysis.

VFA analysis. Analysis of VFAs was performed on rumen fluids following standard methods using high-performance liquid chromatography (HPLC) [22]. In brief, after treatment with

calcium hydroxide and cupric sulfate, followed by precipitation of excess calcium with H_2SO_4, portions of each rumen liquid fraction were run on an HPX-87H HPLC column (BioRad, Hercules, CA) with a mobile phase of 0.015 N H_2SO_4 in 0.0034 M ethylenediaminetetraacetic acid (EDTA) at a flow rate of 0.7 mL/min and a temperature of 45 °C [22] against an external standard containing equimolar concentrations of all C_2-C_5 VFAs plus lactate and succinate. Crotonic acid, added with the $CuSO_4$ treatment, served as an internal standard.

Genomic DNA Extraction. Total genomic DNA was extracted separately from the ruminal solids and liquids following a mechanical disruption and hot/cold phenol extraction protocol published previously [23] with the following modification: 25:24:1 phenol:chloroform:isoamyl alcohol was used in place of phenol:chloroform. This method is essentially identical to the high-yield PCSA method of [24] that results in genomic ruminal bacterial DNA considered to be highly representative of the ruminal bacterial community. After quantification using a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE), DNA pools were made for each cow by combining equal masses of DNA from each cow's three successive daily samples in order to average out minor day-to-day fluctuations in microbial content. All DNA samples were stored at -20 °C.

DNA amplification and sequencing. Each DNA pool was checked for degradation by gel electrophoresis on a 1% agarose TAE gel and by Nanodrop prior to use in PCR. Universal primers were designed to generate amplicons for each sample across the variable V6-V8 regions of the bacterial 16S rRNA gene and included twenty unique 5 bp barcodes on the reverse primer ("XXXXX" in the given sequence) and the adapters A and B suitable for Lib-L Titanium 454 pyrosequencing (926F-5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGAAACTYAAAKGAATTGACGG-3' and

1392R-5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-XXXXX-ACGGGCGGTGTGTRC-3'). A complete set of sample IDs, run names, and barcodes is given with the sequence data deposited in the National Center for Biotechnological Information's Short Read Archive projects under accession SRP027210. A total of 25 ng of DNA and 0.125 μM of each primer was used in a 50 μL reaction including the high-fidelity DNA polymerase Platinum Blue master mix (Invitrogen Life Technologies, Grand Island, NY) with the following cycling conditions: initial denaturation of 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 the final extension at 68 °C for 10 min. Removal of primer and small DNA fragment contaminants was performed by cleaning twice with Agencourt AMPure XP (Beckman Coulter, Inc., San Diego, CA) magnetic beads following manufacturer guidelines. Each sample was quantified using a Qubit® Fluorometer (Invitrogen, San Diego, CA), then pooled to create a single sample at 1×10^9 molecules per μL and diluted to 1×10^6 molecules per μL for use in emPCR at a ratio of 0.8 molecules per emPCR bead. Recovery and sequencing was performed following the manufacturer's guidelines on a Roche 454 GS Junior pyrosequencer with the Lib-L kit and Titanium chemistry.

Sequence processing. Sequence processing was performed using the program mothur v.1.29.2 [25] with default command parameters, unless specified. In brief, all sequences were de-noised (*shhh.flows*, an implementation of AmpliconNoise algorithm [26] and trimmed (pdiffs=2, bdiffs=0, maxhomop=6, minlength=250) prior to alignment against the SILVA 16S rRNA gene reference alignment database [27]. Chimera detection (*chimera.uchime*) was done on a screened version of the alignment (*filter.seqs*) that had been reduced using *unique.seqs* and *pre.cluster* (diffs=2). Determination of operational taxonomic units (OTUs) was performed using the Greengenes database [28] with a confidence level of at least 80 with Cyanobacterial, Eukaryota,

and Archaeal lineages removed. The following analyses and statistical tests were performed within mothur: Good's coverage [29], Morisita-Horn index [30], nonmetric multidimensional scaling (NMDS, iters = 1,000,000) [31], principal component analysis (PCA) [32], Simpson's diversity index [33], UniFrac [34], and unweighted pair group method with arithmetic mean (UPGMA) clustering using average neighbor clustering [35]. The following statistical tests were performed in Primer 6 v.6.1.13: PERMANOVA [36] and SIMPER analysis with Bray-Curtis similarity [37]. The relationship between specific OTUs and feed efficiency was examined using two statistical models. In the first, the relative sequence abundance for 65 specific OTUs (expressed as a percentage of total sequence reads) whose mean abundance across all cows and samples comprised > 0.3% of the total was analyzed using PROC MIXED in SAS v9.2 (SAS Inst. Cary, NC), with feed efficiency group (high or low efficiency) and phase (liquid or solid fraction) as class variables. Significance was declared at $P < 0.05$ using the Tukey procedure. In the second analysis, percent relative sequence abundance of specific OTUs was regressed against gross feed efficiency as a continuous variable for all cows (i.e., containing both the high and low efficiency groups), using PROC REG in SAS. Separate regression analyses were conducted for the liquid and solid phase samples, and significance was declared when model $P > F$ was <0.05. Additional statistical tests include Grubbs' test for outliers [38] and Student's t-test.

Results

Group efficiency and ruminal chemistry. From our initial group of 13 lactating cows, we identified 4 pairs (8 individuals) that exhibited within-pair similarities in parity (lactation cycle, LC), stage of lactation (days in milk, DIM), pregnancy (days carrying calf, DCC), and within-pair differences in energy utilization. Each pair contained a high-efficiency (HE) and low-

efficiency (LE) cow, relative to each other (Table 1). The HE animal net feed efficiency (FE) values ranged from 7.36 – 8.58 MJ/DMI, while the LE animal net FE values were from 6.75 - 7.52 MJ/kg DMI. The FE ranges overlap because of the use of relative, and not absolute, FE classification. In order to increase the robustness of our analysis did not include efficiency as metadata for the comparison of total community composition until after the comparisons were completed, and for our analyses of differences between HE- and LE-classified communities we also performed the same analyses with the classifications randomized. The randomized correlation values were non-significant ($P > 0.05$) (data not shown). Milk yield and composition, ruminal pH, and dry matter intake (DMI) for the eight paired cows are given in Supplementary Table S2. All cows were in mid-to-late lactation (DIM > 152 days at the start of the trial) based on a standard 305-day lactation cycle [39]. Three pairs were in their third LC and one pair (Pair 4) was nearing the end of their second LC. Two of the pairs (pairs 3 and 4) were nearing or past half-way through their pregnancies (based on an average 280-day gestation period [40]).

We used HPLC to determine the ruminal concentrations of the individual VFAs acetate, propionate, butyrate, isobutyrate, valerate, and the co-eluting pair isovalerate and 2-methylbutyrate, as well as the non-volatiles lactate and succinate (Supplementary Table S3). We found significantly greater concentrations of propionate in the LE cows (21.12 ± 1.94 molar % in LE and 18.88 ± 1.90 molar % in HE, $P = 0.009$). Although succinate mean concentrations were statistically significantly higher in the HE cows, the concentrations were very low (0.01 ± 0.02 molar % in LE and 0.04 ± 0.03 molar % in HE, $P = 0.024$) and this difference was not considered to be biologically relevant. Lactate was only detected in one sample from one cow (cow 3231, 0.14 molar %). Similarly, acetate and isobutyrate were significantly higher in HE

Table 1. Paired host data used to assign pairs differing in relative feed efficiency.

Pair	Cow - ID	Net FE MJ/kg DMI	Net energy partitioning (MJ/d)					Net Energy MJ/d	RFE	DCC	LC	DIM
			Gestation[a]	Mastitis[b]	Body[c]	ΔBW[d]	Milk[e]					
	1HE - 3245	8.58	0	5.54	41.34	39.16	4	218.98	High	0	3	149
1	1LE - 3016	6.75	0	1.84	42.69	10.15	4	178.02	Low	0	3	156
	2HE - 3246	7.36	0	5.54	44.08	33.36	1	219.97	High	0	3	232
2	2LE - 3231	7.11	0	0	43.93	29.01	3	186.69	Low	43	3	203
	3HE - 3039	7.70	18.62	3.70	46.33	36.26	8	215.09	High	207	3	288
3	3LE - 3091	7.10	0	1.84	40.68	13.05	4	161.61	Low	137	3	218
	4HE - 3438	7.72	0	1.84	44.28	33.36	7	206.45	High	116	2	244
4	4LE - 3446	7.52	0	1.84	40.32	39.16	97.63	178.96	Low	172	2	251

ΔBW, Change in body weight; DIM, Days in milk; DCC, Days carrying calf; DMI, Dry matter intake; LC, Lactation cycle; RFE, Relative feed efficiency. DIM and DCC given as first day of the sampling period.

[a] Gestation = [(0.00318 × D) – 0.0352]/0.14 Mcal/d × 4.184 MJ/Mcal, where D = DCC between 190 and 279d; assumes calf birth weight of 45 kg [1].

[b] Mastitis = (Dairy Herd Improvement loss class based on somatic cell count (kg milk/d)) × 0.683 Mcal/kg × 4.184 MJ/Mcal [2].

[c] Body = Body maintenance = (0.080 Mcal/d) × (kg BW)$^{0.75}$ × 4.184 MJ/Mcal [1].

[d] ΔBW = ((ΔBW kg/d)/(15d)) × 5.12 Mcal/kg BW × 4.184 MJ/Mcal.

Table S2. Mean rumen pH, dry matter intake, and milk production and composition values with standard deviation of the mean for each cow.

Cow ID	pH N=3	DMI N=9, kg/d	Milk Production N=9, kg/day	Milk composition (%)			
				Butterfat N=18	Protein N=18	Lactose N=18	SNF N=18
3016	6.5 ±0.1	26.4 ±2.3	22.8 ±2.0	3.1 ±0.3	2.9 ±0.1	5.0 ±0.1	8.7 ±0.2
3039	6.6 ±0.2	28.0 ±1.6	16.6 ±0.9	4.2 ±0.2	3.7 ±0.1	5.0 ±0.1	9.7 ±0.1
3091	6.4 ±0.2	22.8 ±1.3	20.1 ±1.5	3.2 ±0.2	2.6 ±0.1	4.8 ±0.1	8.1 ±0.1
3231	6.5 ±0.2	30.9 ±2.4	21.0 ±2.3	4.7 ±1.0	3.6 ±0.1	5.0 ±0.1	9.5 ±0.1
3245	6.5 ±0.1	25.5 ±1.7	24.4 ±1.9	3.3 ±0.5	2.7 ±0.1	4.8 ±0.1	8.3 ±0.2
3246	6.6 ±0.1	25.4 ±1.9	18.4 ±1.3	3.3 ±0.2	3.2 ±0.1	4.8 ±0.1	8.9 ±0.2
3438	6.7 ±0.1	26.7 ±1.0	19.4 ±1.7	4.4 ±0.8	3.2 ±0.1	4.9 ±0.1	9.1 ±0.1
3446	6.5 ±0.1	23.8 ±1.6	16.8 ±1.2	3.4 ±0.3	3.1 ±0.1	5.1 ±0.1	9.1 ±0.2

DMI - Dry matter intake; SNF – Solids, non-fat

Table S3. Volatile fatty acid and succinate molar percentage in ruminal fluid from HE and LE cows.

Acid	HE group (N=4) molar %[a]	LE group (N=4) molar %	Pooled SEM	*P* [b]
Acetate	65.99	64.22	0.69	0.018
Propionate	18.88	21.12	0.78	0.009
Butyrate	10.49	10.21	0.20	0.160
Isobutyrate	1.23	1.09	0.07	0.041
Valerate	1.37	1.50	0.09	0.250
Isovalerate & 2-methylbutyrate [c]	1.99	1.85	0.10	0.140
Succinate	0.04	0.01	0.01	0.024

[a] Results are mean values from three consecutive daily ruminal samples per cow.

[b] $P > t$ from Student's t-test; gray shading highlights those VFAs with $P < 0.05$.

[c] These VFAs co-elute on the HPLC.

cows ($P = 0.018$ and $P = 0.041$, respectively), but the differences were small (< 2 molar %). To test the importance of our chosen variable of relative efficiency on VFA levels we randomized the VFA data by sorting on cow number, splitting the data in half to mimic the HE/LE groups, and found no significance when comparing the two groups ($P > 0.1$). The remaining VFA proportions were not significantly different between the two groups.

HE and LE cows have highly diverse ruminal bacterial communities. In order to determine the composition and structure of the ruminal bacterial community, we performed 454 pyrosequencing of the V6-V8 region of the 16S rRNA gene using extracted DNA from rumen solids and liquids from each cow. We generated a total of 95,386 sequences, 42,449 of which were unique. We also identified 2,432 unique OTUs at 95 % sequence similarity (corresponding roughly to genus-level identifications [41]) (Supplementary Table S4), 763 of which were classifiable to a named genus with at least 80 % confidence. The sequences had an average length of 435 bp, and a minimum of 300 bp. After all quality filters were performed, we obtained an average of 5,962 ± 428 sequences per sample. When combined into HE and LE groups, the distribution of sequences was 24,988 for HE liquids, 23,907 for HE solids, 23,463 for LE liquids, and 23,035 for LE solids. We identified a total of 17 classifiable and one unclassifiable phyla. Based on a comparison of our sequencing metrics (Supplementary Table S4) to previous pyrosequencing studies from the rumens of lactating cows [42], Good's coverage values of 97 % for all samples, the leveling appearance of rarefaction (Supplementary Fig. S1), Chao1 (Supplementary Fig. S2), and ACE collector's curves (Supplementary Fig. S2), each sample was considered to have sufficient sequence coverage.

The diversity within each sample and across the groups was considered high and even. Specifically, each animal was highly diverse according to the inverse Simpson's index (82.28 to

Table S4. Sequence distribution and coverage for each animal.

Cow	Number of sequences	Unique OTUs	Good's coverage	Inverse Simpson's diversity
1HE	12,058	993	0.971	103.90
1LE	11,728	940	0.972	90.15
2HE	12,891	1,010	0.974	116.65
2LE	12,391	1,021	0.971	82.27
3HE	11,895	1,079	0.968	126.59
3LE	11,651	905	0.972	112.60
4HE	12,051	1,025	0.971	112.42
4LE	10,721	997	0.968	134.41
Total	**95,386**	**2,432**[a]		

[a] Most OTUs were shared among multiple animals.

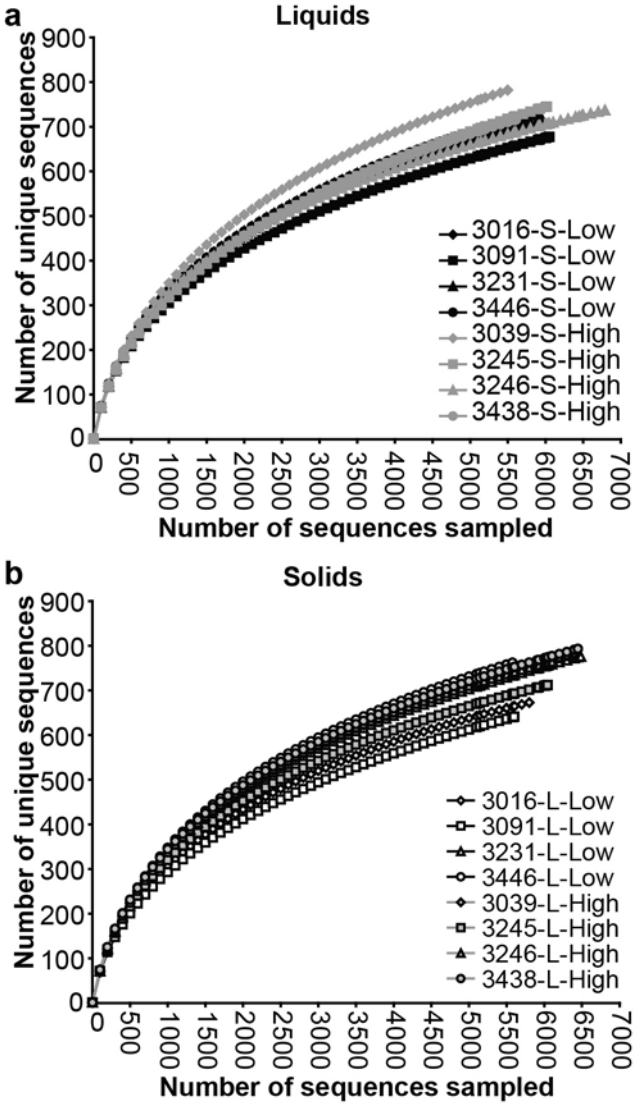

Figure S1. Rarefaction curves for samples from all four pairs of cows based upon the 0.05 OTUs (95% similarity OTU, genus-level) showing completeness of sequence coverage. All samples had 6,338 total sequences (Cow 3446, solid phase) or more, to a maximum of 8,271 total sequences (Cow 3246, solid phase). Cows are identified by: ID number-S(solid) or L(liquid)-Efficiency group (low or high). **(a)** Rarefaction curves for all liquids samples, **(b)** rarefaction curves for all solid samples.

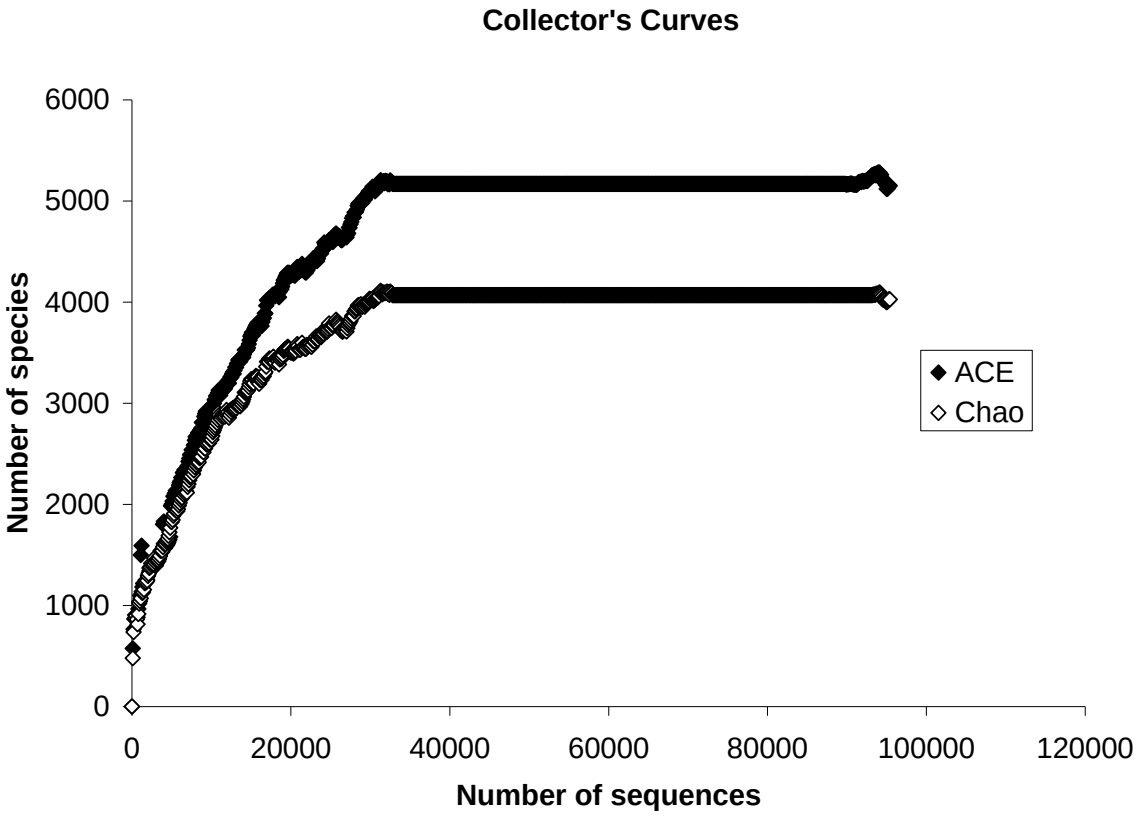

Figure S2. ACE and Chao1 collectors curves for all pooled samples showing completeness of sequence coverage. All sequences were treated as a single sample in order to determine coverage of the rumen bacterial community in the herd.

Table 2. Differences in phylum-level relative abundance for high-efficiency (HE)/low-efficiency (LE) animals and liquid/solid ruminal samples.

	Mean relative sequence abundance (%) [a]				Mean relative sequence abundance (%)			
Phylum	**LE group**	**HE group**	**Pooled SEM**	***P***	**Liquids**	**Solids**	**Pooled SEM**	***P*** [b]
Actinobacteria	0.07	0.07	0.02	0.57	0.04	0.10	0.02	0.002
Armatimonadetes	0.01	0.01	0.01	0.84	0.01	0.01	0.01	0.83
Bacteroidetes	40.12	40.03	4.20	0.98	45.84	33.93	3.08	0.001
Chloroflexi	0.06	0.07	0.03	0.62	0.03	0.11	0.03	0.008
Elusimicrobia	0.01	0.02	0.01	0.29	0.02	BD	0.01	0.003
Fibrobacteres	0.63	0.64	0.13	0.92	0.70	0.57	0.11	0.22
Firmicutes	39.24	40.98	2.87	0.51	31.31	49.31	2.29	<0.0001
Lentisphaerae	0.01	0.03	0.02	0.29	0.01	0.03	0.01	0.061
MVP-15	0.01	BD [c]	0.00	0.16	0.00	0.00	0.00	0.54
Planctomycetes	0.25	0.42	0.09	0.067	0.44	0.23	0.07	0.009
Proteobacteria	4.36	2.56	1.42	0.19	4.11	2.82	1.15	0.25
Spirochaetes	0.89	0.69	0.13	0.14	0.81	0.78	0.12	0.86
Synergistetes	0.07	0.11	0.02	0.099	0.11	0.07	0.02	0.049
Tenericutes	3.96	3.78	0.55	0.72	4.66	3.05	0.42	0.001
TM7	0.15	0.14	0.06	0.87	0.20	0.08	0.04	0.007
Verrucomicrobia	0.14	0.26	0.10	0.21	0.23	0.17	0.08	0.43
WPS-2	0.01	0.02	0.01	0.72	0.01	0.02	0.01	0.60
Unclassified	10.03	10.19	1.96	0.93	11.48	8.72	1.24	0.032

[a] Results are from mean values from four cows within each group (HE and LE).

[b] $P >$ t from Student's t-test; gray shading highlights those VFAs with $P < 0.05$.

[c] BD = Below detection (< 1 sequence per 48,895 sequences in HE; < 1 sequence per 46,942 in solids).

134.41 with an average of 109.87 ± 17.44 SD) (Supplementary Table S4). No sample was significantly lower or higher in diversity ($P > 0.05$, Grubb's test), and when sequences were pooled by group or compared among all individuals there were no significant differences in diversity (Simpson's diversity index) between HE and LE groups ($P = 0.46$, Student's t-test).

Phyla differ between liquids and solids but do not distinguish HE from LE. We identified 18 phyla, of which 5 contributed over 96 % of the total sequences when all samples were pooled: Firmicutes (40.03 %), Bacteroidetes (39.80 %), Unclassified Bacteria (10.10 %), Tenericutes (3.84 %), and Proteobacteria (3.44 %). All pooled LE group, HE group, liquid fraction, and solid fraction percent relative abundance and significance values are given in Table 2. The HE and LE groups were highly similar in phylum distribution, with only two phyla showing any significant differences ($P < 0.10$): HE-associated Planctomyetes and LE-associated Synergistetes. One phylum, MVP-15, was below detection in the HE group. Similar to previous studies that have shown that liquid and solid fractions contain distinct microbial populations due to the relative enrichment for planktonic or fiber-adherent cells, respectively [12,43,44], we found that these fractions were significantly different from one another in our samples ($P = 0.001$, PERMANOVA). The phyla that were most abundant ($P < 0.05$) in the solid fraction than in the liquid fraction were the Actinobacteria, Chloroflexi, and Firmicutes. The phyla that were more abundant ($P < 0.05$) in the liquid fraction than in the solid fraction were the Elusimicrobia (below detection in solids), Bacteroidetes, Planctomycetes, Synergistetes, Tenericutes, and TM7.

HE and LE groups' community composition diverges below phylum level. In order to avoid artificial community clustering we performed the following analyses blind, and only applied metadata (efficiency or fraction identifiers) when labeling the final results. We compared the total community composition for individual solid and liquid samples for each cow using the

Morisita-Horn index represented as an unweighted pair group method with arithmetic mean (UPGMA). This tree showed a separation between HE and LE samples except for the pair 2HE/2LE (Fig. 1A) and a further differentiation between liquid- and solid-associated communities. When analyzed by non-metric multidimensional scaling (NMDS) (Fig. 1B, Yue and Clayton theta, lowest stress = 0.044 and $R^2 = 0.983$) the solid and liquid fractions formed two distinct ($P = 0.001$ by PERMANOVA) and significant groups (across Axis 1 and Axis 3). A significant secondary separation ($P = 0.036$ by PERMANOVA) was found between HE and LE cows (Fig. 1B, across Axis 1 and Axis 2). No link was found between the ruminal fraction and efficiency classification ($P = 0.971$, PERMANOVA). A clustering pattern similar to our whole-community UPGMA tree was observed when analyzed using principal component analysis (PCA) (Fig. 1C), with the bulk of the variation (76.66 %) explained by the first two axes. In all three analyses, cows 2HE and 2LE (Fig. 1) clustered with the opposite community profile from the other high- and low-efficiency samples.

We then included metadata (efficiency or fraction identifiers) as variables when performing the following analyses. We determined the differences between the HE and LE groups using multiple measures: by sequences pooled to genera (Table 3); by a similarity percentage procedure (SIMPER) using OTUs with a Bray-Curtis similarity matrix [37] (Table 3); by individual classifiable OTUs conserved among all animals (Table 3 and Supplementary Table S5); and by individual OTUs correlating by a reduced mixed model (RMM) analysis of the least square mean (Table 4). We also determined the differences between the liquid and solid ruminal fractions using SIMPER (Table 3) and regression analysis.

Using SIMPER with group (HE or LE) identification, the total percentage similarity did not clearly distinguish between HE and LE groups: the average similarity between the HE and

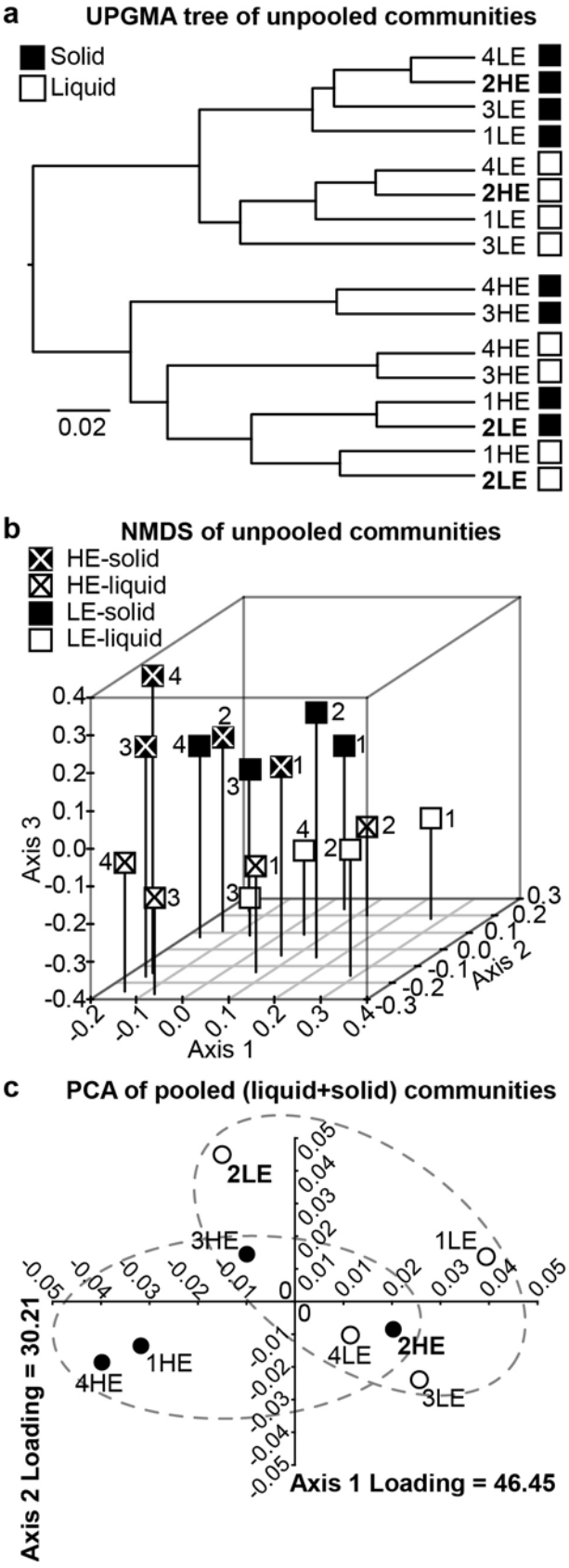

Figure 1. Total community composition is correlated with feed efficiency for high efficiency (HE) and low efficiency (LE) cows. **(a)** UPGMA tree, **(b)** NMDS plot, and **(c)** PCA plot showing total community composition relationships based on 0.05 OTUs and identified by cow ID. In all parts cows are identified by pair number as given in Table 1. In **(a)** and **(b)** filled squares indicate solid, and empty squares liquid, samples from each cow. In **(a)** bolded text indicates the two animals with rumen profiles not matching their designated efficiency classification. In **(b)** exes mark HE cows while empty squares mark LE cows. In **(c)** solid and liquid samples were pooled for each animal, with filled circles representing HE cows, and empty circles LE cows.

Table 3. Distribution of classifiable genera and conserved OTUs by relative sequence abundance and by SIMPER analysis that were significantly different between HE and LE groups.

	Mean relative sequence abundance (%) [a]			
Genus	**LE group**	**HE group**	**Pooled SEM**	***P*** [b]
Lachnospira	0.429	0.297	0.068	0.067
Moryella	0.125	0.192	0.026	0.024
Paludibacter	0.148	0.196	0.025	0.071
Sharpea	0.033	0.007	0.015	0.096
Conserved OTU				
Lachnospira	0.428	0.297	0.068	0.10
Moryella	0.120	0.184	0.023	0.011
Paludibacter	0.133	0.182	0.028	0.062

	SIMPER contributing %			
OTU	**HE vs. LE**	**Association**	**Liquid vs. Solid**	**Association**
Bacteria (100) [c]	3.29	LE	3.34	Liquid
Gammaproteabacteria (100)	3.15	LE	2.55	Liquid
Butyrivibrio (100)	2.18	HE	2.40	Solid
Bacteroidales (100)	1.33	LE	1.61	Liquid
Firmicutes (83)	<1	-	1.33	Solid
Coprococcus (98)	<1	-	1.26	Solid
Lachnospiraceae (100)	<1	-	1.17	Solid
Lachnospiraceae (100)	1.44	HE	1.16	Liquid
Lachnospiraceae (100)	<1	-	1.13	Solid
Lachnospira (100)	<1	-	1.13	Solid
Prevotella (100)	1.04	HE	1.04	Liquid

[a] Results are from mean values from four cows within each group (HE and LE).

[b] $P >$ t from Student's t-test; gray shading highlights those VFAs with $P \leq 0.05$.

[c] Taxonomy for each OTU is given at the highest classifiable level, with the classification confidence in parentheses.

LE groups was 60.45 %, with the average similarity among the HE group of 61.44 % and among the LE group of 62.99 %. The six OTUs with ≥ 1 % dissimilarity for HE vs. LE were among the top OTUs found for liquids vs. solids. For our sequence-based analysis of genera we considered only those genera that contained at least 10 sequences (in order to reduce false inflation of diversity due to singlet sequences) and that were classifiable to the genus level. For our OTU-based analysis of conserved OTUs (Table 3), each OTU had to be present in all eight cows, and be classifiable to the genus level with at least an 80 % confidence. By both measures *Moryella* (608 sequences, 7 OTUs total) and *Paludibacter* (660 sequences, 8 OTUs) were significantly higher in the HE cows, while *Lachnospira* (1,377 sequences in 2 OTUs) was significantly higher in LE cows (Table 3). The genus *Sharpea* (76 sequences in 2 OTUs) was significantly more abundant in LE cows (Table 3) as a whole, but was not present in all animals, being below detection in one LE (3LE) and two HE (3HE and 4HE) cows. There were nine individual OTUs identified by both the SIMPER and by RMM; all such OTUs had the same associations (solid, liquid, HE or LE) by both analyses (Table 4).

At the genus level we found few individual OTUs were responsible for the separation between the solid and liquid phases. By SIMPER (Table 3) the liquid phase was associated with only one classified genus, *Prevotella*, while by regression analysis of OTU abundance against FE we identified four *Prevotella* OTUs to be significant (one at $P < 0.05$ and three at $0.05 < P < 0.01$), and no other OTUs were identified. For the solid phase we identified OTUs in the genera *Butyrivibrio*, *Coporoccocus*, and *Lachnospira* by SIMPER (Table 3), and only two OTUs, a *Prevotella* and an unclassified genus within the Lachnospiraceae, by regression analysis.

Table S5. Distribution of conserved OTUs identified by genus that were not significantly different ($P > 0.10$) by relative sequence abundance between HE and LE groups.

	Mean relative sequence abundance (%) [a]			
Genus	**LE group**	**HE group**	**Pooled SEM**	***P*** [b]
Anaeroplasma	0.074	0.064	0.011	0.54
Bulleidia	0.003	0.004	0.001	0.19
Butyrivibrio	0.694	0.939	0.220	0.19
Coprococcus	0.597	0.579	0.057	0.83
Dehalobacterium	0.005	0.006	0.003	0.55
Eubacterium	0.041	0.059	0.014	0.15
Fibrobacter	0.144	0.152	0.029	0.57
L7A_E11 (Erysipelotrichales)	0.034	0.043	0.011	0.26
Lachnobacterium	0.004	0.004	0.002	0.78
Oscillospira	0.032	0.065	0.025	0.15
p-75-a5 (Erysipelotrichales)	0.046	0.053	0.009	0.27
Prevotella	6.637	6.572	0.765	0.70
Pseudobutyrivibrio	0.267	0.272	0.069	0.77
Pyramidobacter	0.016	0.025	0.006	0.12
RFN20 (Erysipelotrichales)	0.020	0.030	0.010	0.25
Ruminobacter	0.118	0.117	0.048	0.90
Ruminococcus	0.701	0.686	0.077	0.77
Selenomonas	0.168	0.156	0.038	0.93
SHD-231 (Anaerolineales)	0.013	0.017	0.007	0.46
Shuttleworthia	0.230	0.095	0.126	0.28
Succinivibrio	0.140	0.098	0.067	0.56
Sutterella	0.015	0.028	0.010	0.13
Treponema	0.179	0.128	0.030	0.15

[a] Results are mean values for four cows within each group

[b] $P >$ t, Student's t-test

Table 4. Correlation of individual OTUs to efficiency (HE or LE) and phase (Liquids or Solids) by a reduced mixed model (RMM) of relative sequence abundance.

RMM[a] (%) HE	RMM (%) LE	Pooled SEM	*P*	Taxonomy (Confidence)[b]
Higher in HE group				
1.587	0.899	0.162	0.011	*Prevotella* (100)
1.570	0.975	0.169	0.027	*Prevotella* (100)[f, g]
1.528	1.030	0.152	0.038	Ruminococcaceae (100)[e]
1.288	0.814	0.115	0.032	*Prevotella* (100)
1.155	0.748	0.103	0.015	*Prevotella* (100)
0.976	0.560	0.121	0.031	Bacteroidales (100)[c]
0.450	0.301	0.042	0.026	*Prevotella* (100)
Higher in LE group				
1.198	1.715	0.145	0.025	*Lachnospira* (100)[d]
0.836	1.336	0.157	0.043	*Prevotella* (100)
0.624	0.845	0.071	0.048	*Prevotella* (100)
0.208	0.634	0.076	0.002	Lachnospiraceae (100)
RMM (%) Liquids	**RMM (%) Solids**			
Higher in Liquids				
2.802	1.554	0.214	0.001	Bacteroidales (100)[g]
2.293	1.711	0.137	0.010	Bacteroidetes (100)
1.664	0.882	0.169	0.006	*Prevotella* (100)[f, g]
1.293	0.978	0.088	0.025	*Prevotella* (100)
1.213	0.550	0.097	<0.001	*Prevotella* (100)
1.031	0.505	0.121	0.009	Bacteroidales (100)
0.999	0.404	0.184	0.040	*Ruminococcus* (99)
0.900	0.096	0.046	<0.0001	Bacteroidales (100)[c]
0.897	0.119	0.097	<0.0001	*Prevotella* (100)
0.761	0.182	0.135	0.010	*Succinivibrio* (98)
0.694	0.166	0.071	0.0002	*Prevotella* (100)
0.630	0.329	0.069	0.009	Mollicutes-RF39 (100)
0.627	0.404	0.058	0.018	*Prevotella* (100)
0.559	0.231	0.100	0.037	*Prevotella* (100)
Higher in Solids				
1.646	3.711	0.591	0.004	*Butyrivibrio* (100)[g]
0.972	1.941	0.145	<0.001	*Lachnospira* (100)[d, g]
0.722	1.731	0.081	<0.001	*Lachnospiraceae* (100)[g]
1.039	1.519	0.152	0.044	Ruminococcaceae (100)[e]
0.401	1.508	0.089	<0.001	*Coprococcus* (98)[g]
0.307	1.480	0.095	<0.001	Firmicutes (83)[g]
0.698	1.473	0.130	0.001	*Pseudobutyrivibrio* (100)
0.299	1.339	0.179	0.001	Lachnospiraceae (100)[g]

0.492	1.165	0.034	<0.001	*Ruminococcus* (100)
0.406	1.122	0.110	<0.001	Lachnospiraceae (100)
0.354	0.753	0.097	0.012	Lachnospiraceae (100)
0.293	0.601	0.041	0.0001	Lachnospiraceae (99)
0.217	0.550	0.075	0.008	*Ruminococcus* (100)
0.201	0.402	0.055	0.024	Bacteroidales (100)

[a] RMM = Reduced mixed model analysis of the least square mean from four cows within each group (HE or LE), or from eight samples within each phase (liquids or solids).

[b] Taxonomy for each OTU is given at the highest classifiable level, with the classification confidence in parentheses.

[c, d, e, f] Indicate a single OTU significant for both efficiency and phase.

[g] Indicates an OTU also identified to be associated with that condition by SIMPER analysis.

Identification of a shared OTU set and the distribution of classical ruminal bacteria. Among all cows, there was a shared set of 377 OTUs, 158 of which were classifiable to 26 genera. This large number of shared OTUs may have been due in part to the relatively small number of animals (N = 8), and in part to the use of active dairy production animals instead of including animals with production parameters such that they would be normally culled from an active dairy herd.. When classifiable OTUs were pooled by genus, the shared set represented approximately 11 % of all sequences in both the HE (10.85 %) and LE (11.20 %) groups. The most abundant members of this shared set were *Prevotella*, *Butyrivibrio*, *Ruminococcus*, *Coprococcus*, and *Lachnospira* (Supplementary Table S6). Two OTUs were associated with the HE group (*Moryella* and *Paludibacter*) and one with the LE group (*Lachnospira*) (Table 3). When we looked for OTUs present in only the HE or LE groups, there were three OTUs present in all HE cows but absent in all LE cows: an unclassified Ruminococcaceae, an unclassified Clostridia, and an unclassified Bacteria. However, none of the individual HE-unique OTUs were above 0.05 % sequence abundance. There were no OTUs present in all LE cows that were absent in all HE cows.

We then assessed the differences between feed efficiency with respect to the genera *Bacteroides*, *Butyrivibrio*, *Fibrobacter*, *Megasphaera*, *Prevotella*, *Ruminococcus*, *Selenomonas*, and *Streptococcus*. These genera represent many of the most historically studied and physiologically well-characterized ruminal bacterial species [45,46,47,48,49]. We also analyzed their distribution by relative sequence abundance (differences in total sequence count of each genus) and relative OTU diversity (differences in the number of different OTUs within each genus) as shown in Supplementary Table S6. By either metric, we saw no differences between HE and LE groups greater than 1.33-fold, except for *Megasphaera*, which was only detected in

Table S6. Distribution of classical ruminal bacterial genera by relative sequence and OTU abundance.

	Relative sequence abundance (%) [a]				Relative OTU abundance (%) [a]			
Genus	**Mean HE**	**Mean LE**	**Pooled SEM**	**Fold difference (HE/LE)**	**Mean HE**	**Mean LE**	**Pooled SEM**	**Fold difference (HE/LE)**
Bacteroides	0	BD	0	-	0.05	BD	0.03	-
Butyrivibrio	4.09	3.04	0.82	1.35	1.57	1.77	0.14	0.89
Fibrobacter	0.64	0.63	0.12	1.02	0.55	0.59	0.07	0.94
Megasphaera	BD	0	0	-	BD	0.03	0.03	-
Prevotella	28.93	30.29	3.25	0.96	18.74	21.28	1.19	0.88
Ruminococcus	2.79	2.87	0.29	0.97	1.04	1.11	0.06	0.94
Selenomonas	0.63	0.67	0.12	0.93	0.22	0.24	0.04	0.93
Streptococcus	0	BD	0	-	0.03	BD	0.03	-

[a,] Results are mean values from four cows within each group. All genera were not significant ($P > 0.05$, Student's t-test) between LE and HE groups.

BD = Below detection (< 0.002 %, or < 1 sequence in 48,895 for HE and < 1 sequence in 46,491 for LE groups)

the LE group, and *Bacteroides* and *Streptococcus*, which were only detected in the HE group. *Megasphaera* and *Streptococcus* were each represented by a single sequence and *Bacteroides* by two sequences, each with only one OTU. *Prevotella* had the most sequence counts as well as the greatest diversity of OTUs. By relative sequence abundance the most common well-characterized ruminal bacterial genera were *Prevotella*, *Butyrivibrio*, and *Ruminococcus*, with only a minor presence of *Fibrobacter*, *Selenomonas*, *Streptococcus*, and *Megasphaera.*

Discussion

The increasing resolution with which we can characterize microbial communities within a host has led to the realization that complex consortia of bacteria are major drivers of host physiology [50]. Moreover, recent work has demonstrated potential links between the ruminal community, feed conversion efficiency [13], and the production of individual milk components such as milk fat [13,16]. In dairy cows, the ruminal microbial consortium is directly responsible for the conversion of plant matter into volatile fatty acids (VFAs) that are used by the host for metabolism and milk production [51]. We hypothesized that because lactating dairy cows depend on ruminal bacteria for nutrition and milk precursors, there exists a correlation between ruminal bacteria and host performance, specifically in feed efficiency. We tested this hypothesis using a variety of host biometric data with both ruminal VFA analysis and 16S rRNA pyrosequencing in eight dairy cows, classified as either high- or low-efficiency with respect to feed conversion. In general, our findings support our initial hypothesis that milk production parameters and ruminal microbial populations are correlated. Moreover, our work suggests that feed efficiency is correlated with the relative abundance of specific bacterial taxa, although the greatest factor is the total community composition. Our finding that within the same genus (such

as *Prevotella*) specific OTUs associated with either HE or LE animal suggests that species in the same genus may be performing functionally distinct roles in the rumen.

The energy content of milk is strongly influenced by its fat content [1]. Previous work has demonstrated that propionate can depress milk fat levels, while butyrate and acetate can have a stimulatory effect on milk fat levels [14,52,53]. We hypothesized that the concentrations of these VFAs would be significantly different between the HE and LE groups. Our study confirmed that propionate was significantly higher in our LE group, butyrate remained unchanged, and acetate had a modest, but statistically significant, increase in the HE group. However, VFA concentration data must be interpreted cautiously because they do not necessarily mirror the uptake or utilization of VFAs by the host and may be confounded by differences in ruminal liquid phase volume [54].

We found that the total ruminal community included a high abundance of sequences identified as belonging to the phyla Bacteroidetes and Firmicutes. We also found sequences belonging to the Tenericutes, Proteobacteria, and Fibrobacteres along with other numerous phyla of minor abundance. These results are similar to previous work on ruminants [42,43,55,56,57,58], and in agreement with other studies [43,44,57], we found that the planktonic (liquids) and fiber-adherent (solids) bacterial communities were significantly different both at the phylum (Table 2) and lower (Table 4) taxonomic levels. Using a combination of SIMPER and regression analyses we found that the OTUs most responsible for contributing to these differences were often unclassified at the genus level, but included members of *Butyrivibrio*, *Coprococcus*, *Lachnospira*, *Prevotella*, *Pseudobutyrovibrio*, *Ruminococcus*, and *Succinivibrio*. One of the traditionally phase-separable bacteria, the fiber-adherent *Fibrobacteres*, was at too low of a sequence abundance across the dataset to allow for

meaningful discrimination.

Within the constraints of the relatively small number of animals studied, our results also suggest a substantial correlation between the total ruminal bacterial community composition and relative milk production efficiency classification for lactating Holstein cows. Three out of the four cows in both the HE and LE groups had their relative milk production efficiencies predicted by community composition alone (Fig. 1). The degree of correlation among HE and LE animals was reduced due to the presence of a single pair (2HE and 2LE) that had their efficiency classifications opposite to what would have been predicted based on their ruminal community profiles (Fig. 1). This conflicting result could potentially be due in part to the pairs' mismatch in pregnancy status and calculated energy loss from mastitis (Table 1); it is possible that even early pregnancy status perturbs the ruminal bacterial community or energetic calculations or that the energy drain of mastitis interfered with our high/low efficiency classification.

We observed major differences between the HE and LE groups using both traditional taxonomy (Table 3) and sequence identity, OTU-based analyses (SIMPER and regression analyses). By combining all three analyses we identified a number of classifiable genera specifically associated with the HE group: *Moryella*, *Paludibacter*, and *Prevotella*. The possible roles of *Moryella* and *Paludibacter* in the rumen are currently unknown, although *Moryella indoligenes* is weakly saccharolytic and produces acetate, butyrate, and lactate [59]. For the LE group we found associations with *Lachnospira*, *Prevotella*, and *Sharpea*. The physiology of *Sharpea* is largely unknown, while *Lachnospira* is widely regarded as a pectin degrader that produces many fermentation products and whose presence appears to be detrimental to host feed efficiency [60]. Given the presence of multiple *Prevotella* OTUs in both conditions we speculate that this genus includes species with disparate roles in the ruminal community. We further posit

that shifts in the abundance of these community members may play a role in ruminal function and thus feed efficiency, but their actual mechanistic roles must await further physiological characterization.

The shared community members (Table 3 and Supplementary Table S6) found in all animals of our study included many classical ruminally-associated bacterial genera. The high abundance of *Prevotella*, *Butyrivibrio* and *Ruminococcus*, and the minor presence of *Fibrobacter* and *Selenomonas* (Supplementary Table S4), matches previous qPCR-based studies of the rumen [14,23]. Moreover, a comparison of our data set with those of Jami *et al.* [42] and Li *et al.* [61], both of whom used lactating dairy cows and 16S rRNA pyrosequencing, revealed eight genera (*Butyrivibrio*, *Coprococcus*, *Lachnobacterium*, *Moryella*, *Prevotella*, *Pseudobutyrivibrio*, *Ruminococcus*, and *Selenomonas*) present in all animals across all studies. These shared genera may represent essential members of a dairy cow's ruminal bacterial community, as opposed to other historically studied genera (*e.g. Bacteroides* and *Streptococcus*). Finally, we did not find any OTUs unique (that is, present only in) to our LE group, and only three low-abundance OTUs unique to the HE group, suggesting that the presence or absence of individual species within the rumen may not be a predictor of feed efficiency in dairy cows.

In conclusion, we found that the total ruminal bacterial community composition correlates with feed efficiency in dairy cows, with the understanding that other host factors (*e.g.* stage of pregnancy or animal age) and environmental conditions play a role. Our data suggest that there are specific members of the ruminal bacterial community associated with efficiency, including a genus historically associated with the rumen (*Prevotella*) as well as less-studied (*Lachnospira*, *Moryella*, *Paludibacter*, *Sharpea*, and unclassifiable) genera. More detailed associations between individual species and host feed efficiency will require surveys with larger

numbers of animals as well as some physiological characterization of bacterial species associated with HE and LE cows. We acknowledge that the dairy cow is a complicated animal, and that many host and environmental factors are likely to play a role in modifying the ruminal community such as the impact of fetal growth effects [62,63,64]. Based on our findings we speculate that targeted research on genera found to be different between our HE and LE groups could provide new clues in clarifying the ruminal host-microbe relationship. We also recognize that our study does not provide a complete picture of ruminal function, as both fungi and protozoa are known to contribute to the fermentation of plant biomass into VFAs. Moreover, methanogenic Archaea, which are responsible for substantial losses in gross energy, were not examined in this study and should be considered in future work. Our work identifying potentially highly influential ruminal bacteria, in combination with the recent report by Jami *et al.* [16], enables us to create a wide variety of testable hypotheses focused on better elucidating the roles of specific bacteria in influencing final cow milk production efficiency..

Acknowledgements

This work was supported by the University of Wisconsin-Madison College of Agriculture and Life Sciences and a USDA NIFA Fellowship Grant #2012-01193 to K. Jewell. We wish to thank the staff at the USDFRC for their invaluable advice and daily animal care, and in particular N. Betzold for her assistance. We would also like to thank J. Breaker for his support in animal care. We would also like to thank all members of the Suen lab for their support, insightful discussions and careful reading of this manuscript. We also thank M. Hall and C. Arndt (USDA Dairy Forage Research) for helpful discussion on methods for calculating feed efficiency in dairy cattle.

References

1. National Research Council (2001) Nutrient requirements of dairy cattle: National Academy Press, Washington DC.
2. DRMS (2013) The DHI Glossary, August 2013. Raleigh, NC: Dairy Records Management Systems. pp. 20.
3. USDA/NASS (2014) Milk production. Agricultural Statistics Board, USDA: 1-8.
4. USDA/NASS (2014) Agricultural prices. Agricultural Statistics Board, USDA: 1-53.
5. Vlaeminck B, Fievez V, Cabrita ARJ, Fonseca AJM, Dewhurst RJ (2006) Factors affecting odd- and branched-chain fatty acids in milk: A review. Animal Feed Science and Technology 131: 389-417.
6. Fievez V, Colman E, Castro-Montoya JM, Stefanov I, Vlaeminck B (2012) Milk odd- and branched-chain fatty acids as biomarkers of rumen function—An update. Animal Feed Science and Technology 172: 51-65.
7. Guan LL, Nkrumah JD, Basarab JA, Moore SS (2008) Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. FEMS Microbiology Letters 288: 85-91.
8. Deng W, Xi D, Mao H, Wanapat M (2008) The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. Molecular Biology Reports 35: 265-274.
9. Liu S, Bu D, Wang J, Liu L, Liang S, et al. (2011) Effect of incremental levels of fish oil supplementation on specific bacterial populations in bovine ruminal fluid. Journal of Animal Physiology and Animal Nutrition DOI: 10.1111/j.1439-0396.2010.01113.x.
10. Fernando S, Purvis II H, Najar F, Sukiharnikov L, Krehbiel C, et al. (2010) Rumen microbial population dynamics during adaptation to a high-grain diet. Applied and Environmental Microbiology 76: 7482-7490.
11. Tajima K, Aminov R, Nagamine T, Matsui H, Nakamura M, et al. (2001) Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR. Applied and Environmental Microbiology 67: 2766-2774.
12. Petri R, Forster R, Yang W, McKinnon J, McAllister T (2012) Characterization of rumen bacterial diversity and fermentation parameters in concentrate fed cattle with and without forage. Journal of Applied Microbiology 112: 1152-1162.
13. Carberry CA, Kenny DA, Han S, McCabe MS, Waters SM (2012) Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. Applied and Environmental Microbiology 78: 4949-4958.
14. Wang X, Li X, Zhao C, Hu P, Chen H, et al. (2012) The correlation between composition of the bacterial community and concentration of volatile fatty acids in the rumen during the transition period and ketosis in dairy cows. Applied and Environmental Microbiology.
15. Palmonari A, Stevenson DM, Mertens DR, Cruywagen C, Weimer PJ (2010) pH dynamics and bacterial community composition in the rumen of lactating dairy cows. Journal of Dairy Science 93: 279-287.
16. Jami E, White BA, Mizrahi I (2014) Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. PLOS ONE 9: e85423.
17. Weimer PJ, Stevenson DM, Mertens DR, Thomas EE (2008) Effect of monensin feeding and withdrawal on populations of individual bacterial species in the rumen of lactating dairy cows fed high-starch rations. Applied and Environmental Microbiology 80: 135-145.

18. Aikman P, Henning P, Humphries D, Horn C (2011) Rumen pH and fermentation characteristics in dairy cows supplemented with *Megasphaera elsdenii* NCIMB 41125 in early lactation. Journal of Dairy Science 94: 2840-2849.
19. Weimer PJ, Waghorn G, Odt C, Mertens DR (1999) Effect of diet on populations of three species of ruminal cellulolytic bacteria in lactating dairy cows. Journal of Dairy Science 82: 122-134.
20. Tsenkova R, Atanassova S, Toyoda K, Ozaki Y, Itoh K, et al. (1999) Near-infrared spectroscopy for dairy management: measurement of unhomogenized milk composition. Journal of Dairy Science 82: 2344-2351.
21. Connor EE, Hutchison JL, Olson KM, Norman HD (2012) Triennial Lactation Symposium: Opportunities for improving milk production efficiency in dairy cattle. Journal of Animal Science 90: 1687-1694.
22. Weimer PJ, Shi Y, Odt C (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. Applied and Environmental Microbiology 36: 178-183.
23. Stevenson DM, Weimer PJ (2007) Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Applied Microbiology and Biotechnology 75: 165-174.
24. Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, et al. (2013) Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. PLOS ONE 8: e74787.
25. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing MOTHUR: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75: 7537-7541.
26. Quince C, Lanzen A, Davenport R, Turnbaugh P (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12: 38.
27. Pruesse E, Quast C, Knittle K, Fuches B, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35: 7188-7196.
28. DeSantis T, Hugenholtz P, Larsen N, Rojas M, Brodie E, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench campatible with ARB. Applied and Environmental Microbiology 72: 5069-5072.
29. Good I (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40: 237-264.
30. Horn H (1966) Measurement of "overlap" in comparative ecological studies. The American Naturalist 100: 419-424.
31. Borg I, Groenen P (1997) Modern Multidimensional Scaling. New York: Springer.
32. Gower J (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 325-338.
33. Simpson E (1949) Measurement of diversity. Nature 163: 688.
34. Lozupone C, Hamady M, Knight JR (2006) UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 7: 371.
35. Sokal R, Michener C (1958) A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38: 1409-1438.

36. Anderson M (2005) PERMANOVA: a FORTRAN computer program for permutational multivariate analysis of variance. New Zealand: Department of Statistics, University of Auckland.
37. Clarke K (1993) Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18: 117-143.
38. Grubbs FE (1969) Procedures for detecting outlying observations in samples. Technometrics 11: 1-21.
39. Dekkers J, Ten Hag J, Weersink A (1998) Economic aspects of persistency of lactation in dairy cattle. Livestock Production Science 53: 237-252.
40. Knott J (1932) A study of the gestation period of Holstein-Friesian cows. Journal of Dairy Science 15: 87-98.
41. Schloss PD (2010) The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies. PLoS Comput Biol 6: e1000844.
42. Jami E, Mizrahi I (2012) Composition and similarity of bovine rumen microbiota across individual animals. PLOS ONE 7: e33306.
43. de Menezes AB, Lewis E, O'Donovan M, O'Neill BF, Clipson N, et al. (2011) Microbiome analysis of dairy cows fed pasture or total mixed ration diets. FEMS Microbiology Ecology 78: 256-265.
44. Fouts D, Szpakowski S, Purushe J, Torralba M, Waterman R, et al. (2012) Next generation sequencing to define prokaryotic and fungal diversity in the bovine rumen. PLOS ONE 7: e48289.
45. Wallace R, Brammall M (1985) The role of different species of bacteria in the hydrolysis of protein in the rumen. Journal of General Microbiology 131: 821-832.
46. Osborne JM, Dehority BA (1989) Synergism in degradation and utilization of intact forage cellulose, hemicellulose, and pectin by three pure cultures of ruminal bacteria. Applied and Environmental Microbiology 55: 2247-2250.
47. Counotte GHM, Prins RA, Janssen RHAM, deBie MJA (1981) Role of *Megasphaera elsdenii* in the fermentation of dl-[2-^{13}C]lactate in the rumen of dairy cattle. Applied and Environmental Microbiology 42: 649-655.
48. Sijpesteijn AK (1951) On *Ruminococcus flavefaciens*, a cellulose-decomposing bacterium from the rumen of sheep and cattle. Journal of General Microbiology 5: 869-879.
49. Hudson JA, Cai Y, Corner RJ, Morvan B, Joblin KN (2000) Identification and enumeration of oleic acid and linoleic acid hydrating bacteria in the rumen of sheep and cows. Journal of Applied Microbiology 88: 286-292.
50. Nicholson J, Holmes E, Kinross J, Burcelin R, Gibson G, et al. (2012) Host-gut microbiota metabolic interactions. Science 336: 1262-1267.
51. van Houtert MFJ (1993) The production and metabolism of volatile fatty acids by ruminants fed roughages: A review. Animal Feed Science and Technology 43: 189-225.
52. Dijkstra J, Boer H, Van Bruchem J, Bruining M, Tamminga S (1993) Absorption of volatile fatty acids from the rumen of lactating dairy cows as influenced by volatile fatty acid concentration, pH and rumen liquid volume. British Journal of Nutrition 69: 385-396.
53. Ørskov E, Flatt W, Moe P, Munson A (1969) The influence of ruminal infusions of volatile fatty acids on milk yield and composition and on energy utilization by lactating cows. British Journal of Nutrition 23: 443-453.

54. Hall MB (2013) Dietary starch source and protein degradability in diets containing sucrose: Effects on ruminal measures and proposed mechanism for degradable protein effects1. Journal of Dairy Science 96: 7093-7109.
55. Kim M, Morrison M, Yu Z (2011) Status of the phylogenetic diversity census of ruminal microbiomes. FEMS Microbiology Ecology 76: 49-63.
56. Lee J, Jung J, Oh Y, Lee S, Madsen E, et al. (2012) Comparative survey of rumen microbial communities and metabolites across on caprine and three bovine groups, using bar-coded pyrosequencing and ^{1}H nuclear magnetic resonance spectroscopy. Applied and Environmental Microbiology 78: 5983-5993.
57. de Oliveira MNV, Jewell KA, Freitas FS, Benjamin LA, Tótola MR, et al. (2013) Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. Veterinary Microbiology 164: 307-314.
58. Callaway TR, Dowd SE, Edrington TS, Anderson RC, Krueger N, et al. (2010) Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing. Journal of Animal Science 88: 3977-3983.
59. Carlier J-P, K'ouas G, Han XY (2007) *Moryella indoligenes gen. nov*., *sp. nov*., an anaerobic bacterium isolated from clinical specimens. International Journal of Systematic and Evolutionary Microbiology 57: 725-729.
60. Cotta M, Forster R (2006) The family Lachnospiraceae, including the genera *Butyrivibrio*, *Lachnospira* and *Roseburia*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes: Springer US. pp. 1002-1021.
61. Li RW, Wu S, Baldwin VI RL, Li W, Li C (2012) Perturbation dynamics of the rumen microbiota in responses to exogenous butyrate. PLOS ONE 7: e29392.
62. Smith JW, Legates JE (1962) Relation of days open and days dry to lactation milk and fat yields. Journal of Dairy Science 45: 1192-1198.
63. Carman GM (1955) Interrelations of milk production and breeding efficiency in dairy cows. Journal of Animal Science 14: 753-759.
64. Linn J, Trulla M, Casper D, Raeth-Knight M (2004) Feed efficiency of lactating dairy cows. St. Paul, MN: University of Minnesota.

CHAPTER THREE

Ruminal bacterial community composition correlates with gross feed efficiency in lactating Holstein cows over multiple lactations

Kelsea A. Jewell, Caroline McCormick, Christine L. Odt, Paul J. Weimer, and Garret Suen

To be submitted for publication

Milk and rumen samples were collected by KJ, CM, and GS. KJ and CM processed the rumen and feed samples. KJ performed the sequencing reactions, primary sequence analyses, and wrote the manuscript. CO performed the feed composition analysis. CO and PW performed the HPLC analysis of the VFA samples. PW, GS, and KJ designed the experiments. GS assisted with writing the manuscript.

Abstract

A same-aged group of 14 Holstein cows was followed through the course of their first two lactation cycles (2012-2014). During each lactation cycle the ruminal solids and liquids, milk samples and production data, and feed consumption values were collected for Early (76-82 days in milk, or DIM), Middle (151-157 DIM), and Late (251-257 DIM) lactation periods. The ruminal bacterial community for each sample was determined by sequencing the V6-V8 variable region of the 16S rRNA gene using 454 pyrosequencing, with the sequences binned to 95 % operational taxonomic units (OTUs). The gross feed efficiency (GFE) for each cow was calculated by dividing their energy-corrected milk by dry matter intake (ECM/DMI) for each stage of both lactation cycles. The most abundant phyla detected for all cows (in decreasing relative sequence abundance) were the Bacteroidetes, Firmicutes, Proteobacteria, and Tenericutes. The most abundant OTUs shared among all animals were in the genus *Prevotella*, with the other shared community members composed of *Ruminococcus* and *Coprococcus*, among others. The communities of the solid and liquid phases were significantly different, but the communities between the first and second lactation were highly similar. Diversity increased significantly over the course of the two lactations. There was a significant correlation between total community composition and GFE. The ruminal bacterial community was found to be dynamic in terms of membership, diversity, and richness over time, but to also consistently correlate with GFE.

Introduction

It has long been known that ruminants, such as cattle, rely upon a rich and diverse community of ruminal microbes in order to digest feed. A central and open question in ruminal microbiology is the development and stability of this community, since ruminants are born without a functional rumen and must acquire their cellulolytic microbes from the environment [1]. Further, the membership and stability of the ruminal community can have a direct and measurable impact on host function and health [2], especially as cattle require ruminal fermentation products for body maintenance and growth [3], and milk production [4].

Combining assessment of the ruminal microbiota with established measures of host production and efficiency can yield important insights into the effects of specific microbes on host physiology. Measurement of the ruminal community using PCR-denaturing gradient gel electrophoresis with qPCR has demonstrated correlations between cattle weight gain efficiency (calculated as residual feed intake, or RFI) and ruminal community composition and membership [5,6,7]. Recently, the first work using high-throughput sequencing technology to examine ruminal microbes in the context of milk production reported correlations for milk composition, RFI, and individual genera within the ruminal bacterial community [5]. It remains to be determined to what extent the results of this study are applicable to cattle in general (as opposed to effects dependent upon study conditions), and especially to what degree bacterial-host correlations remain stable over time. For dairy cattle, in particular, there are major physiological shifts associated with stage of lactation and pregnancy [9-11], meaning that any conclusions drawn about production efficiency and ruminal community in dairy cows must take into account time as a significant variable.

Investigations into time-dependent changes in the ruminal community have to date focused on the first year of life, as during that time the anatomical structure of the pre-ruminant stomach changes rapidly as the calf's diet shifts from protein- and fat-rich milk to cellulose- and starch-rich feed. Concurrent with anatomical shifts it is thought that the microbial community similarly changes to emphasize digestion of feed over milk, as it has been shown that as calves begin eating feed they concurrently increase ruminal concentrations of microbial fermentation products such as volatile fatty acids (VFAs) and lactate [6]. Recent work using stomach liquids and pyrosequencing of 16S rRNA gene to monitor the bacterial community have shown that colonization of the pre-ruminant stomach may begin within one day after birth [7], although this finding has been disputed [8]. As calves mature there are major and significant changes even at the phylum level due to a general convergence towards an "adult" profile numerically dominated by the phyla Proteobacteria, Bacteroidetes, and Firmicutes [7,8,9], in that rank order for all three studies, although our own work as reported here and in Chapter 2 of this thesis has shown that on the diets used in our research the rank of phyla in the adult Holstein cow rumen is Bacteroidetes, Firmicutes, Proteobacteria, and then Tenericutes. The difficulty of defining when an animal is bodily mature as opposed to sexually mature, coupled to the granularity of sampling (comparing 42 d-old to 1 y-old for [9] and 6 m-old to 2 y-old for [7]) and assumptions regarding community similarity when animals are sacrificed [9] or sampled concurrently instead of sequentially [7], leaves unresolved many questions regarding ruminal community dynamics. When the same set of calves was followed over the course of a single 83 d period it appeared that the community was no longer a function of time after only 15 d [8], possibly because from that point onward the calf diet consisted of at least 15 % hay and starter in addition to milk replacer.

The present study we examined the dynamics of the ruminal bacterial community for

both solid and liquid phases of the rumen contents over the course of two sequential years for the same group of dairy cows. We sampled from each cow during her first and second lactation cycles, thereby being able to compare community composition both across (first lactation *versus* second lactation) and within each lactation (Early, Middle, and Late) for a total of six periods. By sequencing our long-term sample collection we were able to investigate links between the ruminal bacterial community and host production over time at a previously impossible degree of clarity.

Materials and methods

Animal care. A group of 22 Holstein heifers (*Bos taurus*) were selected from birth in the US Dairy Forage Research Center (USDFRC) farm herd (Prairie du Sac, WI). All heifers were in the same age-year (~2 years of age). Cow breeding, veterinary and daily care were performed by USDFRC staff in accordance with farm established protocol and the animal usage guidelines. The heifers were cannulated in early spring of 2012 prior to the start of their first lactation cycle. The heifers were co-housed and, once their first lactation cycle began, kept in tie-stalls in a single barn with other members of the USDFRC milking herd. Once lactation began all cows were milked twice daily by USDFRC staff. The entirety of this study was performed according to RARC protocol A01104, approved by the University of Wisconsin-Madison College of Agriculture and Life Sciences Animal Care and Use Committee.

Feed composition. All animals had *ad libitum* access to water and were fed once daily post morning milking with a total mixed ration (TMR) [10] for the two-year period of the study. The major TMR components were alfalfa haylage, corn silage, and ground corn with toasted soybeans. TMR composition varied slightly due to component availability and price, with every

effort made to maintain a consistent diet for all cows during each lactation cycle. The TMR formulations used during the study are given in Table 1.

Sampling periods. All cows were sampled at three periods during each lactation cycle based on individual days in milk (DIM): 76-82 DIM ("Early"), 151-157 DIM ("Middle"), and 251-257 DIM ("Late") periods based on a standard 305 d lactation cycle [11]. Sample collection began in July and ended in March for both the first (2012-2013) and second (2013-2014) lactation cycles.

Feed and milk sample collection and processing. Feed intake, refusal weights, and feed and refusal samples were collected daily following established methods [12] for each cow over seven days. In brief, sub-samples of TMR and refusals covering the seven-day period were stored at -20 ºC prior to splitting for drying at 105 ºC/24 h for dry matter determination. Milk samples were collected for seven consecutive days starting one day after feed sample collection (as morning milk production would be most impacted by the previous day's feeding) during morning and evening milking by the USDFRC and stored at 4 °C. Milk samples were submitted to AgSource Cooperative Services (Verona, WI) for near-infrared spectroscopic prediction [13] of percent milk fat, lactose, protein, non-fat solids, milk urea nitrogen and somatic cell count. Samples with >7% milk fat were considered spurious and removed from the dataset. Energy corrected milk (ECM) was calculated using true protein as: kg milk × [(0.0929 × % fat) +

Table 1. Feed ingredients for total mixed ration (TMR) diet.

Component	First lactation % DM	Second lactation % DM
Corn silage	32.9	33.7
Alfalfa haylage	26.2	23.3
High-moisture corn, finely ground	14.8	16.6
Dry corn	5.7	3.6
Roasted soybeans	7.4	8.1
Canola meal	6.3	7.7
Distillers dried grains	4.2	4.2
Vitamin/mineral mix	2.5	3.0

(0.0563 × % true protein) + (0.0395 × % lactose) [14]. Ruminal pH was tested during sampling using a portable pH meter or colorimetric pH strips as part of health monitoring.

Rumen sample collection and processing. Ruminal solids and liquids were collected from the bottom of the rumen through the cannula once daily prior to morning feeding (~7:00 am) on three successive days starting four days prior to the end of each sampling period. The mixed solids and liquids were strained through four layers of cheesecloth to collect liquids, followed by additional squeezing to remove remaining liquid from the solids. Solid and liquids samples were kept in sterile 50 mL conical tubes on wet ice and immediately transported back to the lab and frozen at -80 °C until DNA extraction was performed.

Genomic DNA Extraction. Total genomic DNA was extracted separately from the ruminal solids and liquids following a mechanical disruption and hot/cold phenol extraction protocol published previously [15] and similar to the PCSA method of [16], which has been shown to generate high-quality, high-abundance DNA representative of the ruminal bacterial community. In brief, 50 mL of rumen solids were blended with extraction buffer (100 mM Tris/HCl, 10 mM EDTA, 0.15 M NaCl, pH 8.0), centrifuged for 15 min at 500 ×g and filtered through four layers of cheesecloth to remove large particles, and then centrifuged for 1 hr at 5,000 ×g to collect loosened fiber-adherent cells; 50 mL of rumen liquids were centrifuged directly to collect cells. For each collected cell pellet 1 mL was mechanically disrupted with 0.5 g of 0.1 mm zirconia/silica beads (BioSpec Products, Bartlesville, OK) with 50 µL of 20% SDS and 700 µL equilibrated cold phenol on a Mini-beadbeater (Biospec Products) for 2 min, then placed at 60 °C for 10 min, and beaten again. After separating the phases by centrifugation at 4 ºC there followed three phenol extractions and two 25:24:1 phenol:chloroform:isoamyl alcohol extractions, with a final overnight alcohol precipitation of the DNA. DNA was stored in TE at -

80 ºC after quantification on a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE). DNA pools were made for each cow and three day sampling period by combining 2 µg of total DNA from each individual sample into a single tube (for example, all solid-derived DNA for cow 4255 for the Early period of the first lactation). All DNA pools were stored at -20 °C.

DNA amplification and sequencing. We used universal bacterial primers for the 16S rRNA gene covering the variable V6-V8 regions (926 – 1392 bp by the *E. coli* gene numbering system). Our primers included forty unique 5 bp barcodes on the reverse primer ("XXXXX" in the given sequence) and the adapters A and B suitable for Lib-L Titanium 454 pyrosequencing (926F-5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGAAACTYAAAKGAATTGACGG-3' and 1392R-5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-XXXXX-ACGGGCGGTGTGTRC-3'). The de-multiplexed .sff files with sample information and barcodes are deposited with the public National Center for Biotechnological Information's Short Read Archive projects under accession SRP042991. A total of 25 ng of DNA and 0.125 µM of each primer was used in a 50 µL reaction including the high-fidelity DNA polymerase Herculase II Fusion (Agilent Technologies, Santa Clara, CA) with the following cycling conditions: initial denaturation of 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 the final extension at 68 °C for 10 min. Removal of primer and small DNA fragment contaminants was performed by gel extraction in Aqua*Pōr* LM low-melt agarose (National Diagnostics, Atlanta, GA), visualized using SYBR Safe DNA gel stain (Invitrogen, San Diego, CA), and extracted using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA). Each sample was quantified using a Qubit® Fluorometer (Invitrogen), then pooled to create a single sample at 1×10^9 molecules per µL. Any remaining small fragments were removed from the pool using the PureLink PCR Purification Kit (Invitrogen)

with the B3 buffer designed to remove DNA fragments <300 bp. The eluted pool was diluted to 0.8×10^6 molecules per μL for use in emPCR at a ratio of 0.24 molecules per emPCR bead. Recovery and sequencing was performed following the manufacturer's guidelines on a Roche 454 GS Junior pyrosequencer with the Lib-L kit and Titanium chemistry.

Sequence processing and analysis. Sequence processing was performed using the program mothur v.1.33.0 [17] with default command parameters, unless specified. In brief, all sequences were de-noised (*shhh.flows*, an implementation of AmpliconNoise algorithm [18]) and trimmed (pdiffs = 2, bdiffs = 0, maxhomop = 6, minlength = 250) prior to alignment against the SILVA 16S rRNA gene reference alignment database [19]. Chimera detection (*chimera.uchime*) was done on a screened version of the alignment (*filter.seqs*) that had been reduced using *unique.seqs* and *pre.cluster* (diffs=2). Determination of operational taxonomic units (OTUs) was performed using the Greengenes database [20] with a confidence level of at least 80 with Cyanobacteria, Eukaryota, and Archaea lineages removed. Community metrics (coverage, diversity, and richness) and comparisons (principal component analysis, or PCA) were performed in mothur; family-level differences by T-Test on the arcsin-normalized percent relative sequence abundance were performed with PROC TTEST in SAS 9.4 (SAS Inst., Cary, NC), and 1-way ANOVA with Tukey's multiple comparison with Prism 5.0 (GraphPad Software, La Jolla, CA). For comparisons using percent relative sequence abundance, sequences that were not classifiable at the family level, and families with < 0.1 % total relative sequence abundance, were not included.

Results

Cow retention and health. Over the course of both lactation cycles (2012-2014) we retained 14 of our original 22 cows. Cows were not retained if (a) impregnation failed after more than six

artificial insemination attempts, (b) impregnation succeeded but was too delayed to allow for sample collection during the "Early" second lactation window of July to September 2013, or (c) there were health concerns for the animal. During the two years of the trial most cows experienced at least one detected case of mastitis (infection of the udder) and were given antibiotic treatment. Mastitis was defined as a somatic cell count (SCC) of > 300,000 cells/mL in a milk sub-sample, diagnosis by farm personnel, or as *Streptococcus* or *Staphylococcus* spp. cultures recovered from milk. Antibiotic treatments not associated with mastitis were for injuries; all antibiotic treatments were topical and/or local. The rumen sampling periods and associated events are reported in Fig. 1 with cows identified by their farm-originating ID. One cow, ID 4275, experienced a displaced abomasum shortly before her final sampling period (1/5/2014, at 207 DIM) that was surgically corrected. This resulted in a strong negative impact on her milk production (< 2.5 kg/day), with no concurrent drop in feed consumption, for the remainder of the trial.

Sequencing and coverage metrics. We generated 330,811 sequences that passed all filtering metrics, with an average length of 438 bp and a minimum length of 300 bp across all samples. Individual sample sequence counts ranged from 532 to 20,434 sequences (Table 2), with a median of 1,503 sequences and a mean of 1,968 sequences. From all samples we identified 5,793 unique OTUs at 95% sequence similarity (corresponding to genus-level identifications [21]). When pooled by phase (solid plus liquid ruminal fractions) the sample counts ranged from 1,589 to 23,339, with a median of 3,540 and a mean of 4,026 sequences.

Coverage was sufficient for comparison between samples and analysis of major changes in high abundance (defined here as > 0.1 % relative sequence abundance) ruminal bacterial OTUs, and for comparison of total sequence sets for community shifts. The Good's coverage

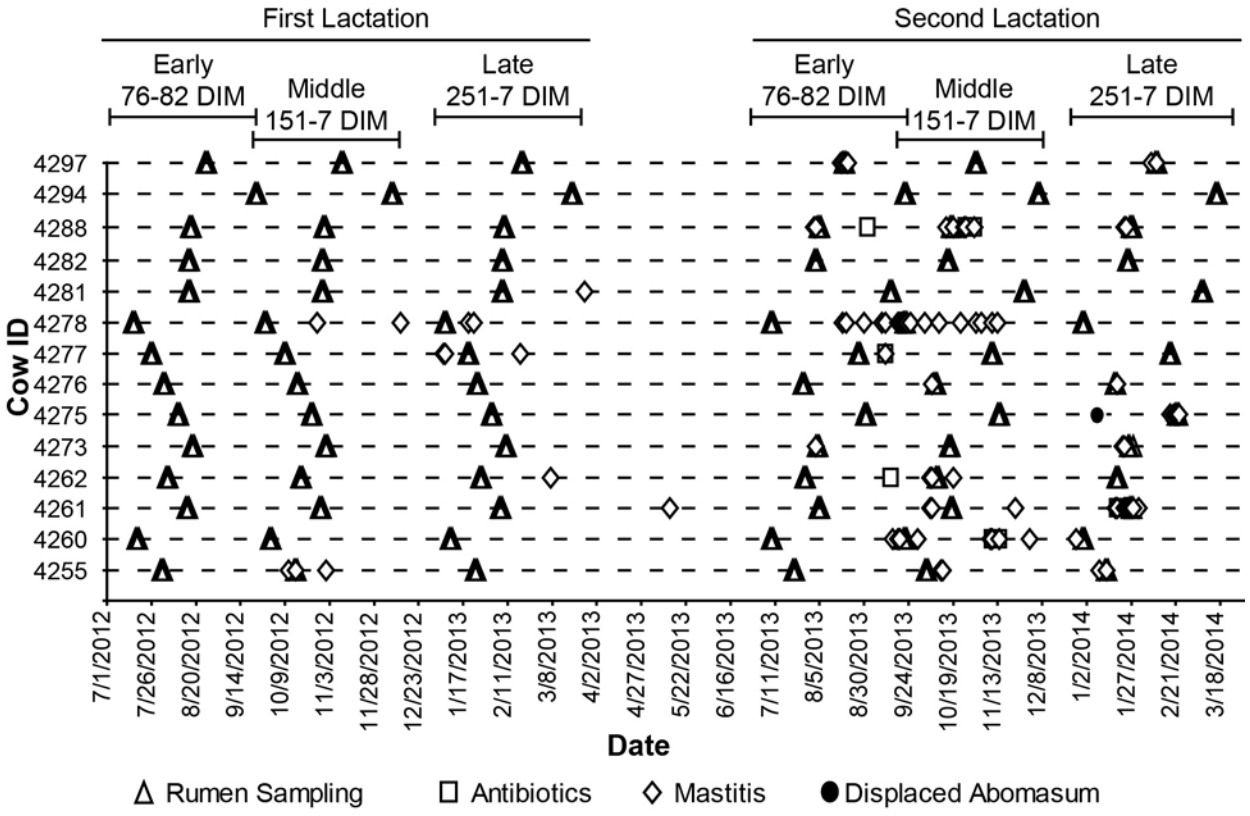

Figure 1. Sampling periods and health-impacting events for the 14 cows retained through both lactation cycles. Triangles mark rumen sampling days (three successive days), rectangles mark topical or localized antibiotic treatments, diamonds indicate mastitis detection (culture- or SCC-based definition), and the single oval indicates a displaced abomasum for cow 4275.

[22] for each sample ranged from 0.77-0.98 for the liquid fraction and from 0.77 - 0.94 for the solid fraction, with a median and mean value of 0.87 across all samples. When the sequence sets were pooled to make a combined solid plus liquid community for each animal the Good's coverage had a mean and median of 0.92, with a minimum of 0.86 and a maximum of 0.98. The mean diversity (Simpson's) of all pooled samples was 99.95, and the median 102.40, with a maximum of 145.78 (cow ID 4261, Second lactation, Early DIM) and a minimum of 36.44 (cow ID 4276, First lactation, Early DIM). The minimum diversity for cow 4276 was not due to incomplete sampling (based on a Good's coverage of 0.94).

General ruminal community composition. The most abundant phyla for all combined samples were the Bacteroidetes (49.42 %), Firmicutes (39.32 %), Proteobacteria (5.67 %), and Tenericutes (2.17 %); the remaining phyla were under 1 % of total sequences. The phyla between 0.1 - 1.0 % relative abundance were the Spirochaetes (0.92 %), Fibrobacteres (0.47 %), TM7 (0.29 %), Planctomycetes (0.27 %), and Verrucomicrobia (0.10 %). Three phyla were unique to the liquids, but were also represented by < 4 sequences each and were considered potentially spurious or transient, low-abundance members of the ruminal community. Unclassified sequences composed 0.96 % of the total.

The genera shared among all animals were *Butyrivibrio*, CF231 (in the Paraprevotellaceae), *Clostridium*, *Coprococcus*, *Fibrobacter*, *Lachnospira*, *Prevotella*, *Pseudobutyrivibrio*, *Ruminococcus*, *Succiniclasticum*, *Succinivibrio*, and YRC22 (in the Paraprevotellaeae). There were 51 individual 95 % OTUs, comprising 44.57 % of all sequences, shared among all animals. Of these OTUs, most were in the genus *Prevotella* (23 OTUs, 25.19 % total relative sequence abundance) with only 1-2 OTUs shared for any other genus. The other shared OTUs were in the genera *Butyrivibrio*, CF231, *Clostridium*, *Coprococcus*, *Lachnospira*,

Table 2. Sequencing, richness, and diversity metrics for all samples.

Cow ID	Cycle	Stage [a]	Phase	Sequence counts	Coverage (Good's)	95% OTUs	Richness (Chao1)	Diversity (Inverse Simpson)
4255	First	Early	Liquid	1,035	0.80	362	707.02	71.37
4255	First	Early	Solid	1,523	0.86	409	797.55	65.74
4255	First	Middle	Liquid	1,296	0.83	417	742.96	117.04
4255	First	Middle	Solid	1,381	0.87	360	585.04	78.22
4255	First	Late	Liquid	778	0.77	322	575.40	157.51
4255	First	Late	Solid	1,558	0.85	442	825.96	109.19
4255	Second	Early	Liquid	1,210	0.88	289	461.08	51.32
4255	Second	Early	Solid	4,079	0.94	560	883.20	55.79
4255	Second	Middle	Liquid	1,252	0.81	423	909.00	127.48
4255	Second	Middle	Solid	3,473	0.93	590	883.04	92.20
4255	Second	Late	Liquid	1,160	0.84	375	602.51	151.09
4255	Second	Late	Solid	566	0.76	225	435.35	72.25
4260	First	Early	Liquid	1,164	0.83	370	673.09	102.37
4260	First	Early	Solid	1,693	0.88	430	708.67	75.20
4260	First	Middle	Liquid	1,795	0.84	512	966.19	100.08
4260	First	Middle	Solid	2,031	0.86	536	938.68	122.36
4260	First	Late	Liquid	1,422	0.84	424	751.08	130.23
4260	First	Late	Solid	1,931	0.88	437	815.42	67.08
4260	Second	Early	Liquid	1,226	0.86	324	576.17	78.66
4260	Second	Early	Solid	3,364	0.94	494	703.01	46.95
4260	Second	Middle	Liquid	1,212	0.82	396	812.27	101.56
4260	Second	Middle	Solid	2,922	0.91	566	908.18	74.66
4260	Second	Late	Liquid	1,360	0.84	412	776.01	90.99
4260	Second	Late	Solid	2,752	0.93	445	648.60	46.78
4261	First	Early	Liquid	2,858	0.92	501	868.97	72.53
4261	First	Early	Solid	1,209	0.86	317	570.61	52.24
4261	First	Middle	Liquid	2,631	0.90	567	972.01	115.16
4261	First	Middle	Solid	1,892	0.88	430	811.82	81.56
4261	First	Late	Liquid	1,643	0.87	435	759.01	101.51
4261	First	Late	Solid	1,711	0.88	410	706.13	105.12
4261	Second	Early	Liquid	1,685	0.86	454	775.16	118.82
4261	Second	Early	Solid	865	0.81	287	531.50	91.34
4261	Second	Middle	Liquid	1,347	0.83	417	757.50	116.21
4261	Second	Middle	Solid	1,241	0.84	355	693.35	79.05
4261	Second	Late	Liquid	1,226	0.82	411	708.69	124.76
4261	Second	Late	Solid	2,294	0.91	439	761.72	65.89
4262	First	Early	Liquid	1,967	0.85	535	1,058.14	122.98
4262	First	Early	Solid	1,600	0.88	399	645.33	92.80
4262	First	Middle	Liquid	1,659	0.86	454	811.00	97.75
4262	First	Middle	Solid	1,251	0.83	377	741.88	102.76
4262	First	Late	Liquid	1,786	0.88	435	760.54	78.99

4262	First	Late	Solid	1,383	0.86	389	653.69	107.81
4262	Second	Early	Liquid	1,274	0.81	438	776.58	164.02
4262	Second	Early	Solid	3,764	0.92	651	1,108.53	99.10
4262	Second	Middle	Liquid	1,535	0.82	477	961.20	111.97
4262	Second	Middle	Solid	1,662	0.87	444	729.80	109.59
4262	Second	Late	Liquid	1,447	0.83	466	851.01	142.36
4262	Second	Late	Solid	2,341	0.91	459	674.97	82.18
4273	First	Early	Liquid	1,355	0.85	356	804.50	56.82
4273	First	Early	Solid	1,206	0.89	282	478.63	46.53
4273	First	Middle	Liquid	1,321	0.86	371	654.52	69.55
4273	First	Middle	Solid	1,975	0.90	425	713.77	68.18
4273	First	Late	Liquid	1,343	0.84	445	685.90	137.69
4273	First	Late	Solid	1,491	0.86	400	719.51	94.12
4273	Second	Early	Liquid	1,249	0.84	372	676.50	77.53
4273	Second	Early	Solid	1,619	0.86	397	789.10	50.14
4273	Second	Middle	Liquid	1,162	0.82	403	688.00	130.90
4273	Second	Middle	Solid	3,680	0.94	564	891.32	77.12
4273	Second	Late	Liquid	1,214	0.85	375	604.40	112.36
4273	Second	Late	Solid	3,231	0.93	490	865.39	59.91
4275	First	Early	Liquid	1,804	0.87	472	783.57	94.69
4275	First	Early	Solid	1,377	0.85	372	747.94	89.43
4275	First	Middle	Liquid	1,523	0.85	433	799.60	86.80
4275	First	Middle	Solid	1,170	0.83	350	728.87	98.53
4275	First	Late	Liquid	1,365	0.84	402	749.84	92.41
4275	First	Late	Solid	1,166	0.87	316	504.91	76.37
4275	Second	Early	Liquid	1,496	0.82	477	933.28	112.24
4275	Second	Early	Solid	2,711	0.92	517	798.28	101.46
4275	Second	Middle	Liquid	1,356	0.84	415	661.17	114.54
4275	Second	Middle	Solid	2,287	0.92	423	691.52	81.91
4275	Second	Late	Liquid	1,239	0.83	398	720.54	134.34
4275	Second	Late	Solid	2,962	0.92	491	812.16	78.36
4276	First	Early	Liquid	1,331	0.90	283	456.94	32.78
4276	First	Early	Solid	1,387	0.91	276	437.88	36.92
4276	First	Middle	Liquid	1,389	0.86	378	679.95	100.70
4276	First	Middle	Solid	1,012	0.84	284	610.15	74.34
4276	First	Late	Liquid	1,274	0.86	348	631.98	100.66
4276	First	Late	Solid	1,227	0.86	332	596.68	89.05
4276	Second	Early	Liquid	1,503	0.88	379	667.47	96.69
4276	Second	Early	Solid	2,319	0.93	391	611.23	42.13
4276	Second	Middle	Liquid	1,223	0.83	374	701.44	117.42
4276	Second	Middle	Solid	2,421	0.93	418	617.51	66.76
4276	Second	Late	Liquid	4,515	0.94	642	961.38	102.57
4276	Second	Late	Solid	2,294	0.92	409	676.99	73.22
4277	First	Early	Liquid	1,366	0.86	359	653.10	62.34
4277	First	Early	Solid	1,503	0.87	382	691.23	80.79
4277	First	Middle	Liquid	1,251	0.86	322	548.25	46.47

4277	First	Middle	Solid	1,660	0.88	413	668.13	88.40
4277	First	Late	Liquid	2,776	0.91	557	873.89	85.02
4277	First	Late	Solid	1,875	0.90	392	677.05	73.76
4277	Second	Early	Liquid	20,434	0.98	1,226	1,692.67	90.24
4277	Second	Early	Solid	2,109	0.90	455	699.64	91.56
4277	Second	Middle	Liquid	2,684	0.92	468	754.00	81.64
4277	Second	Middle	Solid	2,985	0.94	452	653.66	74.52
4277	Second	Late	Liquid	3,398	0.93	536	918.51	63.19
4277	Second	Late	Solid	1,332	0.88	318	534.57	73.09
4278	First	Early	Liquid	1,074	0.83	301	623.20	40.43
4278	First	Early	Solid	1,121	0.87	277	493.20	55.28
4278	First	Middle	Liquid	2,908	0.91	505	875.91	36.01
4278	First	Middle	Solid	1,582	0.87	385	656.62	73.39
4278	First	Late	Liquid	1,182	0.87	270	561.10	33.42
4278	First	Late	Solid	1,239	0.87	309	537.11	62.28
4278	Second	Early	Liquid	2,501	0.92	454	686.65	66.10
4278	Second	Early	Solid	1,522	0.90	333	504.28	41.44
4278	Second	Middle	Liquid	1,502	0.87	373	659.59	57.18
4278	Second	Middle	Solid	2,680	0.92	474	795.04	89.71
4278	Second	Late	Liquid	2,448	0.91	456	770.50	52.61
4278	Second	Late	Solid	3,061	0.94	389	662.65	37.95
4281	First	Early	Liquid	1,148	0.87	280	555.65	58.21
4281	First	Early	Solid	1,289	0.84	358	701.10	79.35
4281	First	Middle	Liquid	1,020	0.82	338	629.61	104.23
4281	First	Middle	Solid	532	0.77	210	460.00	88.56
4281	First	Late	Liquid	1,287	0.84	370	716.55	77.18
4281	First	Late	Solid	1,162	0.85	320	545.33	77.86
4281	Second	Early	Liquid	3,287	0.92	578	914.49	83.24
4281	Second	Early	Solid	2,718	0.90	559	1,014.07	98.66
4281	Second	Middle	Liquid	4,988	0.95	727	970.16	127.21
4281	Second	Middle	Solid	2,792	0.92	483	770.73	80.39
4281	Second	Late	Liquid	6,342	0.95	798	1,142.88	105.97
4281	Second	Late	Solid	2,470	0.93	377	612.49	53.07
4282	First	Early	Liquid	1,036	0.84	311	515.44	89.18
4282	First	Early	Solid	1,465	0.86	412	664.74	97.85
4282	First	Middle	Liquid	1,230	0.86	328	590.89	59.01
4282	First	Middle	Solid	1,310	0.84	380	757.05	86.82
4282	First	Late	Liquid	1,009	0.85	290	505.33	58.07
4282	First	Late	Solid	1,456	0.88	377	613.91	93.27
4282	Second	Early	Liquid	2,306	0.90	524	813.20	108.92
4282	Second	Early	Solid	3,309	0.93	517	838.76	73.01
4282	Second	Middle	Liquid	1,829	0.88	441	724.11	95.35
4282	Second	Middle	Solid	626	0.79	231	423.00	88.16
4282	Second	Late	Liquid	3,543	0.93	600	890.19	100.64
4282	Second	Late	Solid	2,529	0.93	408	619.17	67.25
4288	First	Early	Liquid	922	0.85	259	464.50	56.82

4288	First	Early	Solid	1,449	0.86	389	647.24	92.22
4288	First	Middle	Liquid	1,094	0.82	331	717.12	64.14
4288	First	Middle	Solid	1,008	0.81	341	635.00	92.89
4288	First	Late	Liquid	737	0.80	240	530.38	44.29
4288	First	Late	Solid	1,097	0.81	373	661.12	94.00
4288	Second	Early	Liquid	2,629	0.90	565	855.51	111.27
4288	Second	Early	Solid	2,749	0.91	505	882.92	63.48
4288	Second	Middle	Liquid	1,189	0.85	355	612.36	96.42
4288	Second	Middle	Solid	2,495	0.91	456	772.97	70.60
4288	Second	Late	Liquid	1,986	0.90	449	681.10	114.91
4288	Second	Late	Solid	2,305	0.92	423	625.84	75.17
4294	First	Early	Liquid	856	0.81	283	552.45	82.83
4294	First	Early	Solid	1,328	0.87	368	608.63	106.12
4294	First	Middle	Liquid	865	0.79	310	660.22	111.02
4294	First	Middle	Solid	1,169	0.86	311	534.16	83.23
4294	First	Late	Liquid	828	0.81	266	541.62	71.49
4294	First	Late	Solid	1,115	0.83	336	595.02	89.71
4294	Second	Early	Liquid	2,236	0.89	476	885.45	74.80
4294	Second	Early	Solid	2,512	0.91	461	806.96	94.32
4294	Second	Middle	Liquid	2,850	0.91	574	872.81	101.59
4294	Second	Middle	Solid	2,066	0.91	401	675.66	64.48
4294	Second	Late	Liquid	3,521	0.93	545	962.24	79.46
4294	Second	Late	Solid	2,250	0.93	359	514.30	35.75
4297	First	Early	Liquid	927	0.83	289	518.69	75.11
4297	First	Early	Solid	744	0.78	272	590.07	81.44
4297	First	Middle	Liquid	6,803	0.94	907	1,425.77	80.68
4297	First	Middle	Solid	1,610	0.86	450	776.33	125.96
4297	First	Late	Liquid	1,380	0.85	385	721.38	74.33
4297	First	Late	Solid	1,388	0.85	394	700.01	105.27
4297	Second	Early	Liquid	1,195	0.81	430	782.58	146.13
4297	Second	Early	Solid	2,962	0.91	585	988.53	107.77
4297	Second	Middle	Liquid	2,089	0.91	450	631.01	98.18
4297	Second	Middle	Solid	2,056	0.89	479	752.39	95.87
4297	Second	Late	Liquid	2,312	0.90	499	838.00	109.53
4297	Second	Late	Solid	1,562	0.89	357	624.00	68.12

[a] Early = 76-82 days DIM, Middle = 151-157 DIM, and Late = 251-257 DIM.

Pseudobutyrivibrio, *Ruminococcus*, *Succiniclasticum*, *Succinivibrio*, YRC22 (in the Paraprevotellaceae), and 15 OTUs unclassified at the genus level.

Changes in the bacterial community as functions of phase and chronology. The sequence sets derived from the liquid and solid ruminal phases were significantly different ($P < 0.0001$, unpaired T-test, two-tailed) and visually separable by principal component analysis (PCA) [23] across the y-axis (Fig. 2). This result is consistent with recent large-scale studies [29-31] of bovine ruminal communities that also showed strong separation between the liquid and solid ruminal communities. We quantified differences between the community composition of the phases using a T-test on the arcsin-normalized percent relative abundance of sequences pooled to the family level for all samples within a lactation cycle. We found that 15 families were significantly different ($P < 0.05$) between the two phases (Table 3), with the liquid phase being highly enriched (> 2-fold) for the bacteria F16 in the phylum TM7, Pirellulaceae in Planctomycetes, and RF16 and S24-7 in Bacteroidetes. Only one family, Christensenellaceae in the phylum Firmicutes, was highly enriched in the solid phase.

A difference in richness (Chao1, Table 2) was detected for the total (pooled first and second lactation cycle sequences) sequence sets ($P = 0.016$), where richness was higher in the liquids (95 % confidence interval: 717.8 - 789.7) than the solids (95 % confidence interval: 655.0 - 726.9). There was increased diversity (Inverse Simpson's, Table 2) in the liquid (95 % confidence interval: 96.32-110.5) over the solid (95 % confidence interval: 66.53 - 80.67) phase for the second lactation cycle ($P < 0.0001$) but not the first lactation cycle ($P = 0.62$). The increased diversity (based on the total number of detected OTUs) in the ruminal liquid over the solid phase could be explained by the selection for a conserved group of fiber-adherent, cellulolytic bacteria [24], or by decreased recovery of bacterial species from solid samples,

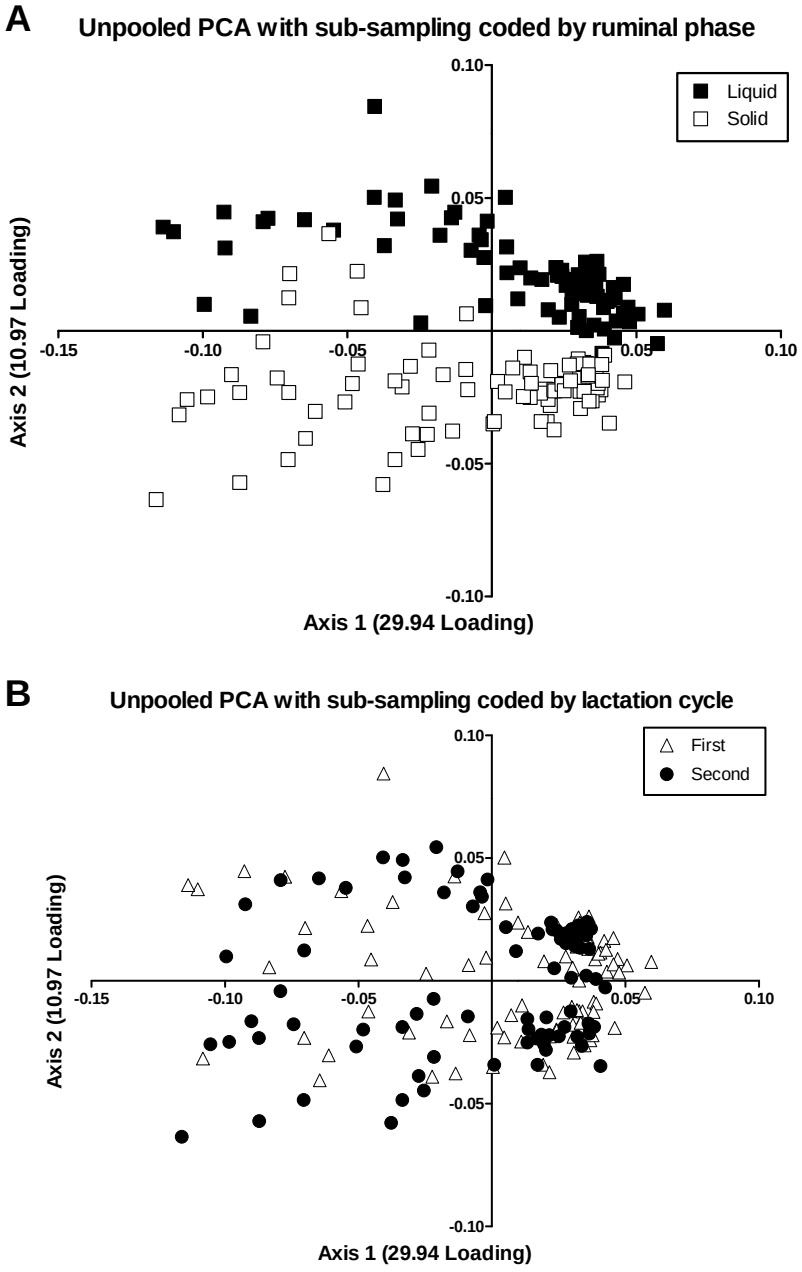

Figure 2. The ruminal communities separate by phase but not by lactation cycle. (A) Black squares = liquid-phase communities. White squares = solid-phase communities. (B) White triangles = first-lactation communities. Black circles = second-lactation communities. Both parts are graphed from the same sub-sampled PCA of all samples.

especially from easily lysed cells such as the Fibrobacteres [16], due to the necessary usage of physical disruption in order to remove cells from feed particles [25].

When compared within the course of the two lactation cycles (Early, Middle, and Late sampling periods) there was a small but steady increase in the diversity for the ruminal liquids (Fig. 3A) with no concurrent change in richness (Fig. 3B). The ruminal solids had an uneven decrease over the two lactation cycles in diversity (Fig. 3C), but within each lactation cycle there is a pattern of decreasing richness from the Early to Late collection periods (Fig. 3D). The total degree of change in diversity or richness in the ruminal community over time, when solids and liquids were pooled, was not significant.

Total community composition of the ruminal bacterial community during the course of two successive lactation cycles. There was little difference by PCA plot between (Fig. 4A) or within (Fig. 4B) comparisons of total community composition for the two lactation cycles, although there was a general area of increased sample density indicating a dominant profile common to many of the cows. A total of ten families were significantly different between the first and second lactation cycles (Table 3), though these differences were modest in value (< 2-fold).

Because we observed changes in microbial diversity and community composition over time, we wished to know if there was a progressive change in community membership common among the cows. To answer this question we calculated vectors based on the *x,y* coordinates of the PCA plot (Fig. 4), setting each cow's start position (First lactation, Early) to (0,0). We found that the total community composition for most cows stays within a narrow range of (0,0), moving no more than 0.04 units away (Fig. 5), although there was a general shift towards the negative *x*-axis (leftward in Fig. 5) over the course of each lactation cycle. The starting community of each lactation cycle (First lactation, Early and Second lactation, Early) did not

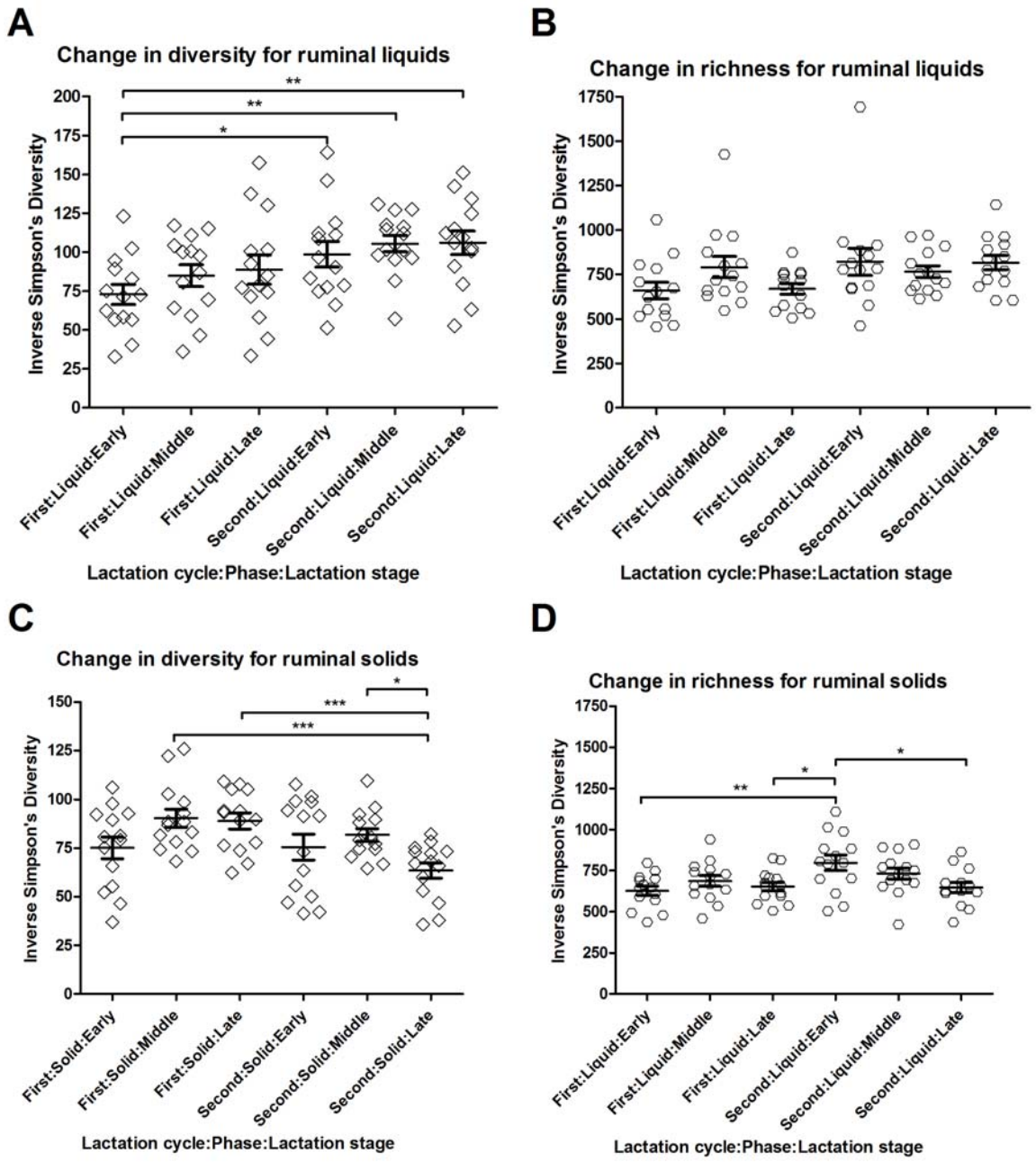

Figure 3. Changes in diversity and richness by phase, lactation cycle, and lactation stage. Correlation of (A) diversity in ruminal liquids, (B) richness in ruminal liquids, (C) diversity in ruminal solids, and (D richness in ruminal solids over time. Early = 76 - 82 DIM, Middle = 151 - 157 DIM, and Late = 251 - 257 DIM. $0.01 > P < 0.05$ (*), $0.001 < P < 0.01$ (**) and $P < 0.001$ (***) by repeated measures 1-way ANOVA with Tukey's multiple comparison.

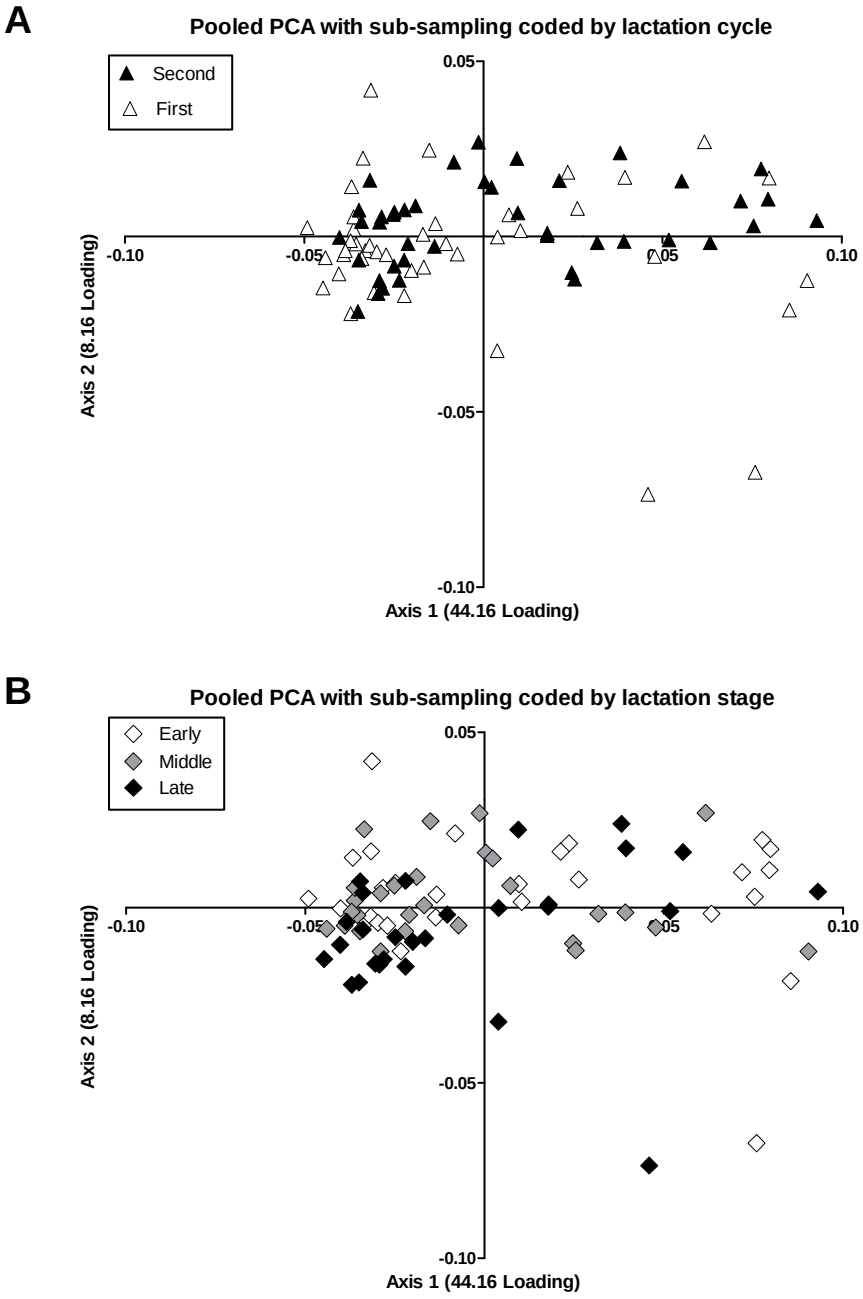

Figure 4. The pooled (solid and liquid) ruminal communities overlap by PCA with sub-sampling. (A) PCA coded by first (empty triangle) or second (filled triangle) lactation cycle. (B) PCA coded by Early (76 - 82 DIM, empty diamond), Middle (151 - 157 DIM, gray diamond), and Late (251 - 257 DIM, black diamond) stages within both lactation cycles.

Table 3. Family-level differences between sequence sets by lactation cycle and ruminal phase as determined by T-Test.

	Relative sequence abundance (%)					
Family	**First lactation**	**SDM [a]**	**Second lactation**	**SDM**	**Equality of variance**	***P* [b]**
BS11 (Bacteroidales) [c]	0.46	0.37	0.31	0.22	0.0012	0.0211
Erysipelotrichaceae	0.93	0.41	1.32	0.85	<0.0001	0.0096
F16 (TM7)	0.39	0.18	0.29	0.11	0.0021	0.0015
Fibrobacteraceae	0.30	0.15	0.54	0.28	0.0001	<0.0001
Lachnospiraceae	15.51	2.87	16.75	3.21	0.4671	0.0665
Mogibacteriaceae (Clostridiales)	0.65	0.20	0.84	0.27	0.0589	0.0003
Paraprevotellaceae	3.20	0.69	3.87	0.77	0.4919	<0.0001
Pirellulaceae	0.34	0.18	0.25	0.14	0.0847	0.0069
Prevotellaceae	43.49	5.59	40.31	4.20	0.0531	0.0039
	Liquids	**SDM**	**Solids**	**SDM**	**Equality of variance**	***P***
Anaeroplasmataceae	0.42	0.29	0.23	0.18	<0.0001	<0.0001
Christensenellaceae	0.08	0.09	0.19	0.13	0.0022	<0.0001
Clostridiaceae	0.94	0.42	2.47	0.62	0.0009	<0.0001
Erysipelotrichaceae	1.59	1.14	0.72	0.32	<0.0001	<0.0001
F16 (TM7)	0.51	0.30	0.20	0.13	<0.0001	<0.0001
Fibrobacteraceae	0.48	0.37	0.38	0.28	0.0076	0.0004
Lachnospiraceae	10.64	3.40	20.68	3.60	0.6045	<0.0001
Mogibacteriaceae (Clostridiales)	0.70	0.32	0.85	0.35	0.0926	0.0066
Pirellulaceae	0.44	0.30	0.18	0.14	<0.0001	<0.0001
Prevotellaceae	45.21	9.93	38.33	6.25	<0.0001	<0.0001
RF16 (Bacteroidales)	0.61	0.35	0.17	0.12	<0.0001	<0.0001
S24-7 (Bacteroidales)	2.69	1.09	0.99	0.54	<0.0001	<0.0001
Spirochaetaceae	0.58	0.35	1.05	0.52	0.0005	<0.0001
Succinivibrionaceae	3.83	2.63	6.07	3.05	0.1834	<0.0001
Veillonellaceae	4.37	1.32	3.12	0.86	<0.0001	<0.0001

[a] SDM = Standard deviation of the mean

[b] When equality of variance was > 0.05 the more conservative Satterthwaite, in place of the Pooled, *P* value is reported.

[c] Additional taxonomic information is given for clarity.

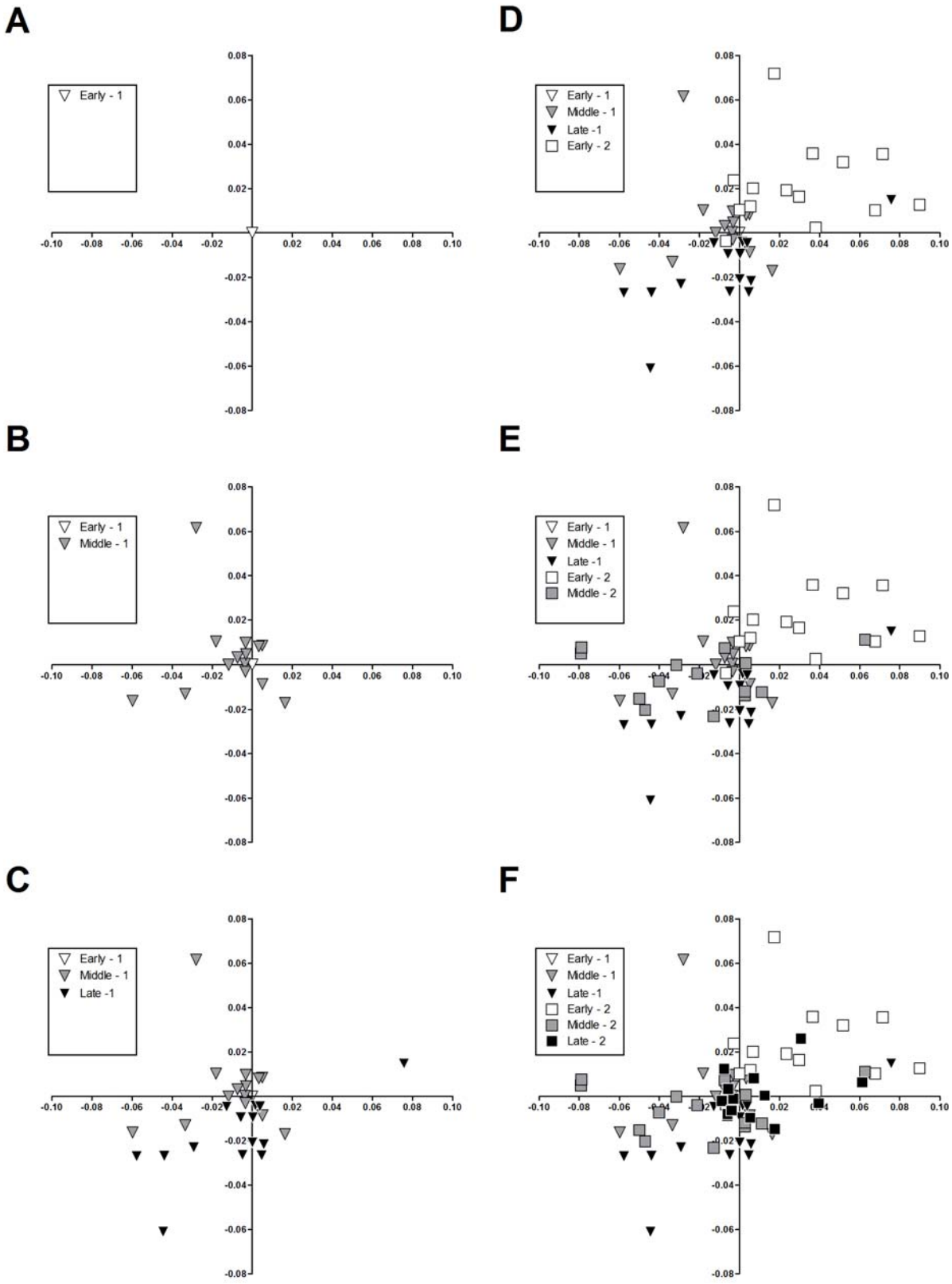
A
Early - 1
B
Early - 1
Middle - 1
C
Early - 1
Middle - 1
Late -1
D
Early - 1
Middle - 1
Late -1
Early - 2
E
Early - 1
Middle - 1
Late -1
Early - 2
Middle - 2
F
Early - 1
Middle - 1
Late -1
Early - 2
Middle - 2
Late - 2

Figure 5. Relative community change over time using PCA coordinates within the pooled, sub-sampled ruminal bacterial communities. (A-F) Step-wise addition of the first lactation and second lactation cycle whole community points as calculated from the relative Early/First PCA point of each cow as (0,0). Triangles indicate all first lactation points; squares indicate all second lactation points. Symbols for early days in milk (DIM) are hollow, while middle DIM symbols are gray and late DIM symbols are black..

overlap (Fig. 4C). The cows that experienced shifts in community composition outside of the 0.04 radius were 4260, 4273, 4276, and 4278 in the first lactation and cows 4255, 4262, 4276, 4278, and 4294 in the second lactation. These cows are not easily differentiated from the more clustered animals (T-test) by community richness ($P = 0.61$), GFE ($P = 0.073$), or community diversity ($P = 0.07$). When the lactation cycles were treated as continuous, rather than discrete, events for the purposes of calculating the degree of change between ruminal community compositions over the course of the entire trial (Fig. 5) it can be seen that the Early points cluster separately from the Middle ($0.01 > P < 0.05$ by 1-way ANOVA with Tukey's multiple comparison) and Late ($P < 0.001$) points, while the Middle and Late points overlap ($P > 0.05$).

Correlation of the ruminal community with gross feed efficiency. Gross feed efficiency (GFE) is defined as energy corrected milk divided by dry matter intake (ECM/DMI) (Table 4). The Second lactation/Late samples for cow 4275 were not included because her displaced abomasum artificially lowered her GFE. We used GFE as a continuous variable, which took into account the shifting relative GFE experienced by all cows over the course of both lactation cycles (Table 5). We measured the degree of correlation between GFE and total community composition by 1-way ANOVA with Tukey's multiple comparison of the linear regression of the first PCA axis coordinates (loading = 44.16, Fig. 4) against ECM/DMI for both the pooled (Fig. 6) and separated ruminal phases (Fig. 7). GFE formed a measurably linear relationship with the Late stage (P=0.04) of the first lactation (Fig. 6A), and the Early stage (P=0.04) of the second lactation (Fig. 6B). When the lactation stages were combined a significant correlation was maintained for the first (P=0.01) lactation; the second lactation had a correlation that was not statistically significant (P=0.11) (Fig. 6C). There was no significant relationship of GFE with diversity (P=0.13, Inverse Simpson's) (Fig. 6D) or richness (P=0.36, Choa1) (Fig. 6E). The

relationships found were not due to differences in sequence coverage (P=0.14, Good's) (Fig. 6F). Thus, we conclude that the correlations represent a *bona fide* relationship between specific members of the microbial community and efficient milk production by the cow.

A linear regression of the GFE against the relative sequence abundance of each classified family identified six families that correlated with increases or decreases in efficiency and had > 0.1 % total relative sequence abundance. Of these, there was a positive correlation with increased GFE for two families: F16 (P = 0.021) and Lachnospiraceae (P = 0.047). There was a negative correlation for four families: Christensenellaceae (P = 0.003), Fibrobacteraceae (P = 0.011), Paraprevotellaceae (P = 0.010), and Ruminococcaceae (P = 0.046).

We measured correlation at the level of individual OTUs by taking the pooled solid and liquid sequence sets for each cow and determining the relative sequence abundance for each OTU. Then, we removed all OTUs that had a total relative sequence abundance of < 0.1 % of the entire sequence set. The range of total sequence abundance per OTU was then 0.10 – 2.15 %, with a mean of 0.39 % and a median of 0.20 % for 158 OTUs. We then performed linear regression on each OTU against the GFE and report those OTUs $P < 0.05$ in Table 6. There were 17 OTUs positively correlated with GFE, of which five were in *Prevotella*, and 32 OTUs negatively correlated with GFE, 15 of which were in *Prevotella*. The largest positive correlation slope was for an OTU in the Succinivibrionaceae, while the largest negative correlation slope was for an OTU in *Prevotella*. There were multiple OTUs identified to Clostridiales, Lachnospiraceae, and *Prevotella* correlating with either increasing or decreasing GFE. OTUs within F16 (Bacteroidales), RF39 (Mollicutes), *Ruminococcus*, *Shuttleworthia*, and Succinivibrionaceae correlated only with increasing GFE. OTUs with *Anaerovibrio*, Bacteroidales, *Butyrivibrio*, CF231 (Paraprevotellaceae), Clostridia, *Fibrobacter*, *Oscillospira*,

Table 4. Energy corrected milk, dry matter intake, and gross feed efficiency values.

	1st lactation								
	Early[a]	Middle	Late	Early	Middle	Late	Early	Middle	Late
	ECM[b]	ECM	ECM	DMI[c]	DMI	DMI	ECM	ECM	ECM
Cow	kg/d	kg/d	kg/d	kg/d	kg/d	kg/d	DMI	DMI	DMI
4255	28.1	28.3	28.4	19.4	23.2	25.9	1.45	1.22	1.10
4260	28.9	27.7	37.1	21.9	24.2	29.0	1.32	1.14	1.28
4261	24.9	26.6	25.9	20.4	22.6	26.5	1.22	1.18	0.98
4262	25.0	26.7	25.8	19.0	23.3	26.3	1.32	1.15	0.98
4273	27.0	27.4	19.5	18.2	21.9	20.8	1.48	1.25	0.94
4275	24.5	26.2	29.6	20.2	23.8	27.7	1.22	1.10	1.07
4276	17.5	22.7	21.8	13.1	19.2	19.7	1.33	1.18	1.11
4277	22.6	26.3	29.3	19.4	27.1	28.1	1.17	0.97	1.04
4278	24.0	21.5	25.4	16.8	17.6	22.2	1.43	1.22	1.14
4281	25.6	31.2	29.0	19.7	24.0	26.6	1.30	1.30	1.09
4282	24.8	25.0	26.5	19.7	23.4	27.5	1.26	1.07	0.96
4288	23.5	24.6	25.7	16.3	22.5	24.8	1.44	1.09	1.04
4294	26.8	27.9	23.9	17.5	19.3	20.0	1.54	1.45	1.19
4297	24.9	28.5	31.2	17.9	25.0	28.7	1.39	1.14	1.09
	2nd lactation								
4255	30.2	35.5	28.2	19.5	29.3	27.7	1.55	1.21	1.02
4260	41.9	34.8	27.8	33.4	30.7	31.0	1.26	1.13	0.90
4261	40.2	39.0	20.6	31.2	33.5	31.8	1.29	1.16	0.65
4262	35.9	40.0	35.4	27.2	30.1	25.3	1.32	1.33	1.40
4273	34.3	35.7	30.0	25.8	25.9	25.3	1.33	1.38	1.19
4275	27.3	27.8	5.70	23.4	29.6	24.4	1.17	0.94	0.23
4276	28.1	31.1	28.6	20.5	22.1	28.2	1.37	1.41	1.01
4277	27.8	26.7	27.0	25.9	33.0	29.0	1.08	0.81	0.93
4278	32.6	25.4	21.6	28.1	27.5	19.7	1.16	0.92	1.10
4281	35.3	33.6	27.3	27.8	26.9	28.2	1.27	1.25	0.97
4282	34.4	32.7	28.9	25.9	30.3	29.8	1.33	1.08	0.97
4288	36.2	37.1	27.1	29.1	27.1	25.1	1.25	1.37	1.08
4294	25.0	31.7	25.2	19.7	28.2	23.3	1.27	1.13	1.08
4297	39.5	37.4	27.2	32.5	31.5	32.2	1.22	1.19	0.85

[a] Stage in lactation is Early (76-82 DIM), Middle (151-157 DIM), or Late (251-257 DIM)

[b] ECM=Energy corrected milk, ECM = kg milk × [(0.0929 × % fat) + (0.0563 × % true protein) + (0.0395 × % lactose)

[c] DMI=Dry matter intake

Table 5. Changes in relative gross feed efficiency (GFE) for all cows over both lactation cycles.

First lactation						Second lactation					
Early		Middle		Late		Early		Middle		Late	
Cow	RGFE [a] %	Cow	RGFE %	Cow	RGFE %	Cow	RGFE %	Cow	RGFE %	Cow	RGFE %
4255	105.3	4255	103.9	4255	103	4255	121.4	4255	104.1	4255	100.7
4260	96.2	4260	97.5	4260	120	4260	98.5	4260	97.2	4260	89.0
4261	89.1	4261	100.3	4261	91	4261	101.1	4261	99.8	4261	64.1
4262	95.8	4262	97.8	4262	92	4262	103.6	4262	114.2	4262	138.4
4273	107.7	4273	106.9	4273	88	4273	104.4	4273	118.6	4273	117.6
4275	88.5	4275	93.8	4275	100	4275	91.6	4275	80.7	4275	NA [b]
4276	97.1	4276	100.9	4276	104	4276	107.3	4276	120.7	4276	100.3
4277	85.0	4277	82.9	4277	98	4277	84.4	4277	69.5	4277	92.3
4278	103.9	4278	104.3	4278	107	4278	91.1	4278	79.2	4278	108.8
4281	94.4	4281	111.1	4281	102	4281	99.7	4281	107.1	4281	95.7
4282	91.7	4282	91.0	4282	90	4282	104.4	4282	92.6	4282	95.8
4288	105.0	4288	93.4	4288	97	4288	97.7	4288	117.6	4288	106.8
4294	111.8	4294	123.5	4294	111	4294	99.4	4294	96.7	4294	106.8
4297	101.3	4297	97.4	4297	102	4297	95.4	4297	101.8	4297	83.6

[a] RGFE = Relative GFE, calculated by dividing each GFE by the mean GFE value for that lactation period such that 100 % = mean GFE value.

[b] 4275 did not recover milk production following surgical correction of a displaced abomasum.

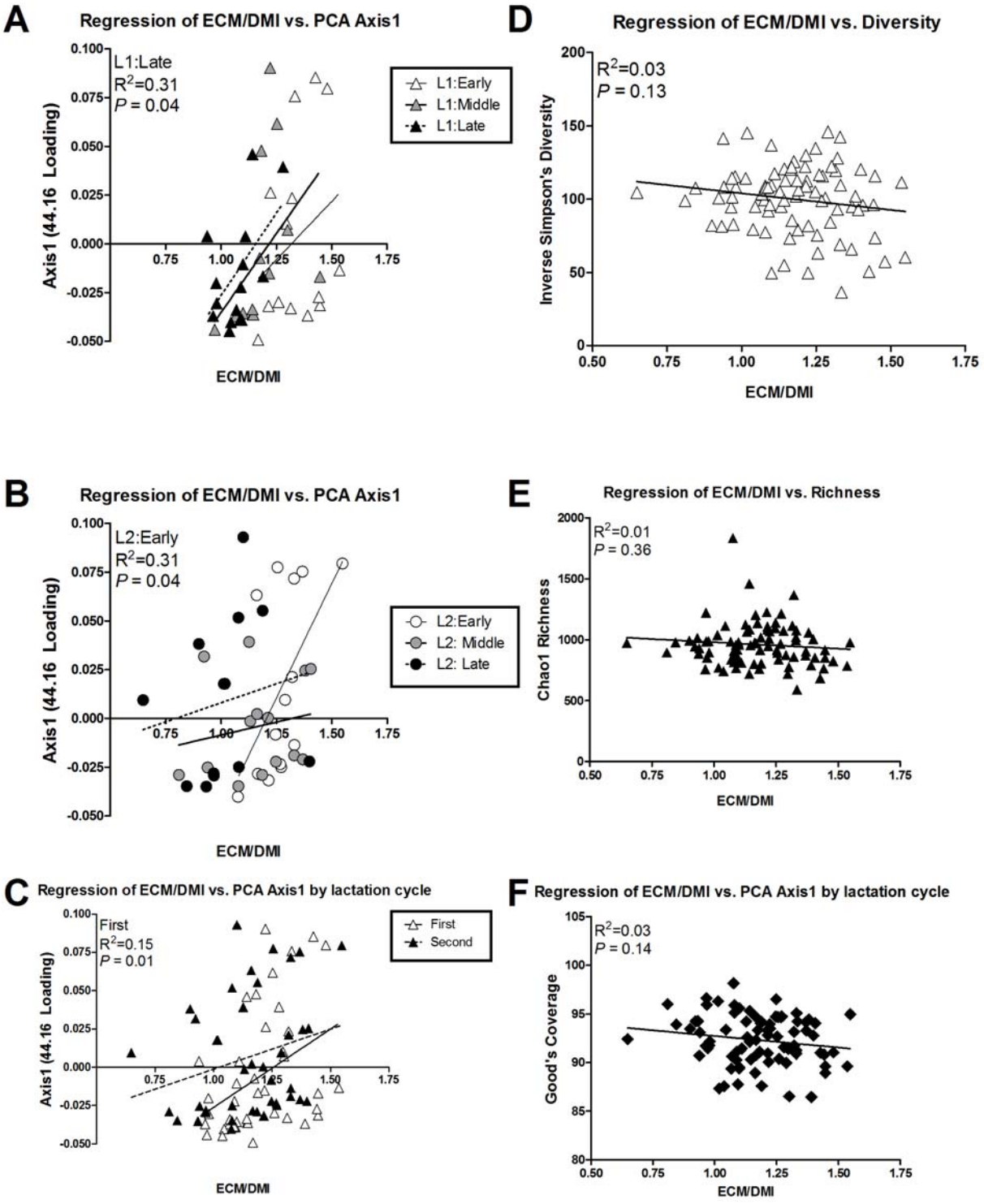

Figure 6. Regression analysis of the gross feed efficiency (ECM/DMI) against the pooled solid- and liquid-phase community, diversity, richness, and coverage. R^2 and *P* values calculated with 1-way ANOVA with Tukey's multiple comparison for (A – B), and by linear regression for (C –

F). L1 = First lactation, L2 = Second lactation, stage in lactation is Early (76 - 82 DIM), Middle (151 - 157 DIM), or Late (251 - 257 DIM). (A - B) First and second lactation with separate stages, $P > 0.05$ not shown. (C) First and second lactation with stages pooled. (D - F) Community diversity, richness, and coverage.

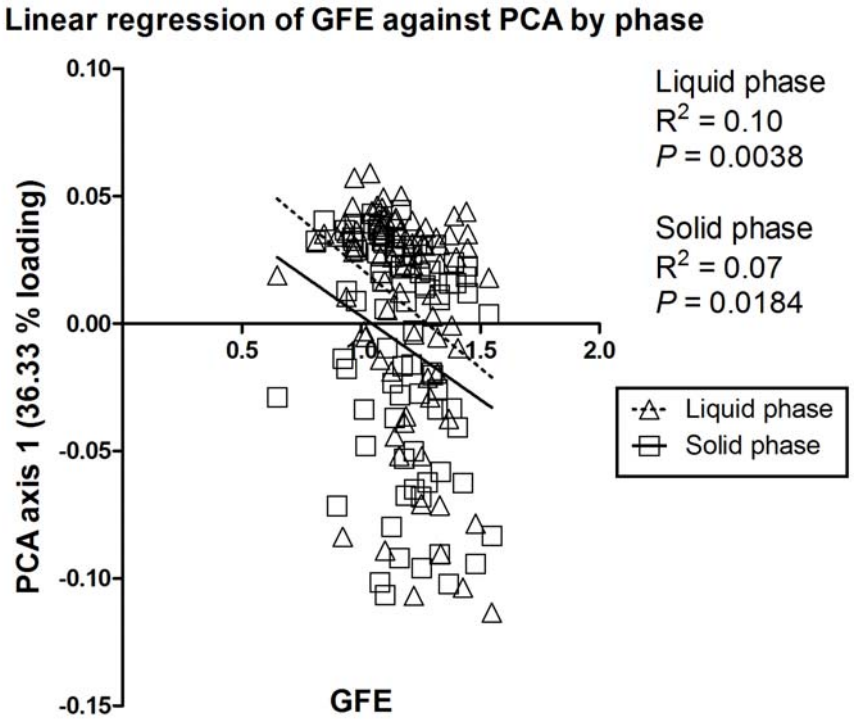

Figure 7. Regression analysis of the gross feed efficiency (ECM/DMI) against the separated solid- and liquid-phase communities. R^2 and *P* values calculated with linear regression. First and second lactation cycles, and periods within each lactation, have been combined.

Table 6. Individual 95 % (genus-level) OTUs with total relative abundance > 0.1 % that correlated by linear regression against gross feed efficiency (GFE)

	Increases with increasing GFE [a]		
OTU	**Slope**	**R^2**	***P***
Succinivibrionaceae [b]	0.03900	0.0935	0.0047
Lachnospiraceae	0.02649	0.0951	0.0043
Prevotella	0.01872	0.0474	0.0466
Shuttleworthia	0.01152	0.0696	0.0153
Prevotella	0.00884	0.0650	0.0193
Clostridiales	0.00437	0.1151	0.0016
Prevotella	0.00349	0.0586	0.0265
Prevotella	0.00340	0.0755	0.0114
Lachnospiraceae	0.00334	0.0469	0.0478
Shuttleworthia	0.00294	0.0542	0.0332
Ruminococcus	0.00287	0.0921	0.0050
Clostridiales	0.00276	0.0681	0.0165
Prevotella	0.00243	0.0996	0.0035
RF39 (Mollicutes) [c]	0.00182	0.0607	0.0238
Lachnospiraceae	0.00178	0.0631	0.0212
F16 (Bacteroidales)	0.00135	0.0830	0.0079
Lachnospiraceae	0.00129	0.0722	0.0135
	Increases with decreasing GFE		
OTU	**Slope**	**R^2**	***P***
Prevotella	-0.00125	0.0628	0.0215
Prevotella	-0.00131	0.0514	0.0382
Ruminococcaceae	-0.00141	0.0531	0.0350
Prevotella	-0.00143	0.0780	0.0101
Treponema	-0.00149	0.0503	0.0404
Clostridium	-0.00151	0.0655	0.0188
Lachnospiraceae	-0.00171	0.0683	0.0164
Oscillospira	-0.00175	0.0936	0.0046
Anaerovibrio	-0.00186	0.0520	0.0369
Prevotella	-0.00208	0.0719	0.0137
Prevotella	-0.00246	0.0893	0.0058
CF231 (Paraprevotellaceae)	-0.00257	0.0675	0.0170
Prevotella	-0.00301	0.0507	0.0395
CF231 (Paraprevotellaceae)	-0.00324	0.1855	<0.0001
Fibrobacter	-0.00330	0.0993	0.0035
Prevotella	-0.00345	0.0981	0.0037
Clostridia	-0.00353	0.1266	0.0009
Clostridiales	-0.00366	0.1678	0.0001
Clostridiales	-0.00395	0.0521	0.0368
Bacteroidales	-0.00399	0.0793	0.0095

Prevotella	-0.00455	0.0537	0.0340
Prevotella	-0.00463	0.0626	0.0217
Prevotellaceae	-0.00468	0.1134	0.0017
Prevotella	-0.00470	0.0701	0.0149
Bacteroidales	-0.00570	0.0728	0.0131
Lachnospiraceae	-0.00571	0.1550	0.0002
Butyrivibrio	-0.00800	0.0506	0.0396
Prevotella	-0.00803	0.0562	0.0299
Prevotella	-0.00834	0.0852	0.0071
Prevotella	-0.00843	0.0712	0.0141
Prevotella	-0.01312	0.0741	0.0123
Prevotella	-0.01397	0.1502	0.0003

[a] GFE ranged from 0.648 to 1.548, excluding the GFE value for 4275 of 0.234 due to recovery from surgically correcting a displaced abomasum (Second, Late).

[b] Each OTU is reported with the nearest classifiable taxonomic level.

[c] Additional taxonomic information is given for clarity.

Prevotellaceae, Ruminococcaceae, and *Treponema* correlated only with decreasing GFE.

Discussion

The physical limitations of maintaining and performing research on the same group of animals over a long time course, coupled with the lack of technology capable of measuring the complexity of the ruminal bacterial community, have restricted the number and scope large-scale microbial surveys. Further, research focused on the development of the ruminal community over time has relied primarily upon calves [13-15] and thus could not include data on efficiency of body weight gain or milk production. The only large-scale sequencing study published to date comparing the ruminal community to milk production parameters took all ruminal samples on a single day [5], and thus cannot be used to address community stability or long-term relationship to host metrics. The present study is novel in that it explicitly addresses the problematic task of assessing the relationship between ruminal microbes and host metrics over an extended period of time. We tracked the ruminal bacterial community over the course of multiple years, with three time points within each of two sequential lactation cycles, for 14 dairy cows, all of which were maintained on the same feed and housed in the same barn. We used 454 pyrosequencing of the V6-V8 region of the 16S rRNA gene to create whole-community profiles for each cow and time point, allowing us to examine changes in the ruminal community as a function of host parameters.

We found that there is a correlation between the total ruminal bacterial community and the gross feed efficiency (GFE), validating the findings reported in Chapter 2 of this thesis. GFE, unlike the high/low efficiency (HE/LE) classifications used in Chapter 2, does not include terms for maintenance, bodyweight change, gestation or mastitis, and is a continuous instead of a

discrete classification. In the work presented in this chapter all cows were of the same parity and gestation stage, and were still actively growing during their first lactation making it difficult to separate changes in bodyweight due to lactation or growth. Expanding the terms of the GFE calculations to include mastitis and bodyweight terms, as well as the inclusion and analysis of the volatile fatty acid presence and concentration in the rumen at each sampling date, was not within the scope of this thesis.

The shared ruminal community that we detected was similar at the phylum level to previous pyrosequencing-based studies of cattle ruminal bacteria showing that the most abundant phyla are Firmicutes, Bacteroidetes, and Proteobacteria [5,26,27]. The only previously published study relating the bacterial community to host production [5] proposed that the ratio of Firmicutes to Bacteroidetes was predictive of milk fat production as they found statistical correlation between that ratio and higher/lower milk fat percentages. However, in our own work we found that there was no correlation ($P = 0.45$, $R^2 = 0.01$, regression analysis) between the ratio of these two phyla and GFE, suggesting that even if this ratio is linked to milk fat it is not predictive of total production efficiency. Our results also may differ due to differences between the two studies such as feed composition, cow breed, age, and housing conditions.

One reasonable expectation would be that, once established, the ruminal community would be stable except in cases of major feed or host health perturbations. Work with calves leveraging the power of large-scale 16S rRNA gene sequencing has shown that bacterial colonization of the pre-ruminant stomach begins within days after birth, with the community converging on an “adult” profile as the calves age, and in particular as they are fed increasing percentages of dietary roughage [8]. The idea of a stable adult ruminal community with only minor, small-scale fluctuations was supported by a study using whole community fingerprinting

(capillary electrophoresis–single strand conformation polymorphism) of adult cows over the course of 21 days [28].

Our results show that the ruminal community cannot be assumed to be stable in the adult cow between, or during, lactation cycles. For the entire group of cows there was a slow but steady increase in diversity of the bacterial community from the ruminal liquids, and a slight decrease in diversity for the ruminal solids, as the cows aged (Fig. 3). The increase in liquids-associated diversity may be due to the continued establishment of low-abundance bacterial populations as the cows age, with the stability of the solid-associated community diversity minimizing changes in total community composition over time (Fig. 4).

For four of our cows in the first lactation period, and for five cows in the second lactation period, there was a high degree of dissimilarity (Fig. 5) between the total ruminal community at each sampling period (Early, Middle, and Late). In addition, the Early communities clustered significantly (Fig. 5) apart from the Middle and Late communities. Given that all of the cows in this study were co-housed, fed the same rations, and sampled at the same DIM, we draw the conclusion that the ruminal community cannot be assumed to be stable over time. Further, there is a conserved shift in all cows away from their individual Early communities as lactation proceeds, and then a return to the Early community profile as the cow enters the next lactation cycle. As our measured host parameters (milk and feeding values) and the sequence set metrics (coverage, diversity, and richness) did not correlate with the degree of shift in the ruminal community it is likely that a host factor, such as the physiological impact of lactation stage, is behind the apparent community instability.

Given that the ruminal bacterial community associated with the solid phase has been shown to be more stable than that associated with the liquid phase in previous work [35-36] and

in this study, it can be conjectured that changes to the solid-associated community are of potentially great importance in terms of feed fermentation. There were strong linear correlations with both the liquid and solid phase communities against GFE (Fig. 7), suggesting that the observed increases in the liquid-associated diversity over time are important influences on feed digestion and host performance. The total increases in diversity over time did not significantly correlate with GFE (Fig. 6D), although there was a slight inverse trend, and so we conclude that changes in the population of specific ruminal bacteria have the greatest impact on GFE. We found multiple ruminal community members significantly associated with both phase (Table 3) and GFE (Table 6). For example, in the solid phase the Lachnospiraceae were significantly correlated with increased GFE; the Lachnospiraceae includes many genera of known rumen function such as *Lachnospira* and *Butyrivibrio* [29]. Conversely, the Prevotellaceae were associated with the liquid phase, and individual members of that family in the well-known ruminal genus *Prevotella* were found with both positive and negative correlations against GFE.

Our regression analysis of individual OTUs revealed that multiple OTUs within Clostridiales, Lachnospiraceae, and *Prevotella* correlated with increasing or decreasing GFE, which may be due to the rise and fall closely related, but more or less beneficial, species. The genera with OTUs only correlating to increasing GFE were *Ruminococcus* and *Shuttleworthia*. The genera with OTUs only correlating to decreasing GFE were *Anaerovibrio*, *Butyrivibrio*, *Clostridium*, *Fibrobacter*, *Oscillospira*, and *Treponema*. There is no obvious pattern by substrate utilization or fermentation products among these two groups of genera based on known isolate characteristics [29,30,31].

There is a strong correlation between the total ruminal bacterial community and the GFE of each cow (Fig. 6) that was not due to community diversity or richness. This result validates

the result of an earlier study from our group (see Chapter 2 of this thesis) that showed, using cows paired for relative efficiency, general clustering of the total bacterial community by high or low efficiency. That this pattern has now been demonstrated twice, with two separate cohorts of cows and over a total of three years of sampling dates, strongly supports the conclusion that the ruminal community plays a measurable and critical role in shaping the efficiency of milk production in the dairy cow. We expect that future studies including other members of the ruminal community, such as the fungi, protozoa, and methanogens, will further elucidate and elaborate upon this conclusion.

Acknowledgements

This work was supported by the University of Wisconsin-Madison College of Agriculture and Life Sciences and a USDA NIFA Fellowship Grant #2012-01193 to K. Jewell. We wish to thank the staff at the USDFRC for their invaluable advice and daily animal care, particularly herd managers Nancy Betzold and Ron Skoyen, and Kurt Pickar and Jacob Breaker for their support in animal care. We would also like to thank all members of the Suen lab for their support, insightful discussions and careful reading of this manuscript. We thank G. Brennan (DVM) for helpful discussions on cattle health. We also thank M. Hall and C. Arndt (USDA Dairy Forage Research) for helpful discussion on methods for calculating feed efficiency in dairy cattle.

References

1. Fonty G, Gouet P, Nebout JM (1989) Development of the cellulolytic microflora in the rumen of lambs transferred into sterile isolators a few days after birth. Canadian Journal of Microbiology 35: 416-422.
2. Wang X, Li X, Zhao C, Hu P, Chen H, et al. (2012) The correlation between composition of the bacterial community and concentration of volatile fatty acids in the rumen during the transition period and ketosis in dairy cows. Applied and Environmental Microbiology.
3. van Houtert MFJ (1993) The production and metabolism of volatile fatty acids by ruminants fed roughages: A review. Animal Feed Science and Technology 43: 189-225.
4. Bauman DE, Harvatine KJ, Lock AL (2011) Nutrigenomics, Rumen-Derived Bioactive Fatty Acids, and the Regulation of Milk Fat Synthesis. Annual Review of Nutrition 31: 299-319.
5. Jami E, White BA, Mizrahi I (2014) Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. PLOS ONE 9: e85423.
6. Anderson K, Nagaraja T, Morrill JL (1987) Ruminal metabolic development in calves weaned conventionally or early. Journal of Dairy Science 70: 1000-1005.
7. Jami E, Israel A, Kotser A, Mizrahi I (2013) Exploring the bovine rumen bacterial community from birth to adulthood. ISME Journal 7: 1069-1079.
8. Rey M, Enjalbert F, Combes S, Cauquil L, Bouchez O, et al. (2014) Establishment of ruminal bacterial community in dairy calves from birth to weaning is sequential. Journal of Applied Microbiology 116: 245-257.
9. Li RW, Connor EE, Li C, Baldwin VI RL, Sparks ME (2012) Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. Environmental Microbiology 14: 129-139.
10. Weimer PJ, Stevenson DM, Mertens DR, Thomas EE (2008) Effect of monensin feeding and withdrawal on populations of individual bacterial species in the rumen of lactating dairy cows fed high-starch rations. Applied and Environmental Microbiology 80: 135-145.
11. Dekkers J, Ten Hag J, Weersink A (1998) Economic aspects of persistency of lactation in dairy cattle. Livestock Production Science 53: 237-252.
12. Weimer PJ, Waghorn G, Odt C, Mertens DR (1999) Effect of diet on populations of three species of ruminal cellulolytic bacteria in lactating dairy cows. Journal of Dairy Science 82: 122-134.
13. Tsenkova R, Atanassova S, Toyoda K, Ozaki Y, Itoh K, et al. (1999) Near-infrared spectroscopy for dairy management: measurement of unhomogenized milk composition. Journal of Dairy Science 82: 2344-2351.
14. National Research Council (2001) Nutrient requirements of dairy cattle: National Academy Press, Washington DC.
15. Stevenson DM, Weimer PJ (2007) Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Applied Microbiology and Biotechnology 75: 165-174.
16. Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, et al. (2013) Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. PLOS ONE 8: e74787.
17. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing MOTHUR: Open-source, platform-independent, community-supported software for

describing and comparing microbial communities. Applied and Environmental Microbiology 75: 7537-7541.
18. Quince C, Lanzen A, Davenport R, Turnbaugh P (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12: 38.
19. Pruesse E, Quast C, Knittle K, Fuches B, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35: 7188-7196.
20. DeSantis T, Hugenholtz P, Larsen N, Rojas M, Brodie E, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench campatible with ARB. Applied and Environmental Microbiology 72: 5069-5072.
21. Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Applied and Environmental Microbiology 71: 1501-1506.
22. Good I (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40: 237-264.
23. Gower J (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 325-338.
24. Brulc JM, Antonopoulos DA, Berg Miller ME, Wilson MK, Yannarell AC, et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proceedings of the National Academy of Sciences 106: 1948-1953.
25. Yu Z, Morrison M (2004) Improved extraction of PCR-quality community DNA from digesta and fecal samples. BioTechniques 36: 808-812.
26. Jami E, Mizrahi I (2012) Composition and similarity of bovine rumen microbiota across individual animals. PLOS ONE 7: e33306.
27. Li M, Penner GB, Hernandez-Sanabria E, Oba M, Guan LL (2009) Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen. Journal of Applied Microbiology 107: 1924-1934.
28. Michelland RJ, Monteils V, Combes S, Cauquil L, Gidenne T, et al. (2011) Changes over time in the bacterial communities associated with fluid and food particles and the ruminal parameters in the bovine rumen before and after a dietary change. Canadian Journal of Microbiology 57: 629-637.
29. Russell JB, Rychlik J (2001) Factors that alter rumen microbial ecology. Science 292: 1119-1122.
30. Binek M, Szynkiewicz ZM (1984) Physiological properties and classification of strains of *Treponema* sp. isolated from pigs in poland. Comparative Immunology, Microbiology and Infectious Diseases 7: 141-148.
31. Downes J, Munson MA, Radford DR, Spratt DA, Wade WG (2002) *Shuttleworthia satelles gen. nov., sp. nov.,* isolated from the human oral cavity. International Journal of Systematic and Evolutionary Microbiology 52: 1469-1475.

CHAPTER FOUR

An analysis of the ruminal bacterial microbiota in West African Dwarf sheep fed grass and tree-based diets

Originally published as:

Omoniyi, Lukman A.[§], Kelsea A. Jewell[§], Olubukola A. Isah, Anthony P. Neumann, Chryss F.I. Onwuka, Okanlawon M. Onagbesan, and Garret Suen. (2014) An analysis of the ruminal bacterial microbiota in West African Dwarf sheep fed grass and tree-based diets. *Journal of Applied Microbiology*. 116(5): 1094-1105.

[§]Contributed equally to this work

Modifications have been made for formatting consistency within this thesis, and to correct certain language usage errors. OL, CO, and OO collected the animal samples. OL, KJ, and AP processed the rumen samples and performed the sequencing reactions. KJ performed the data analysis. OL and KJ wrote the manuscript. OI and GS designed the experiments and assisted with writing the manuscript.

Published with permission from The Society for Applied Microbiology and the *Journal of Applied Microbiology*.

Abstract

We measured the impact of supplementing a forage diet with tree-based browse on the ruminal bacterial communities of Nigerian West African Dwarf (WAD) sheep. Fifteen WAD sheep were fed a control diet of forage (*Panicum maximum*), with 12 animals shifted in groups of three to one of four browse-supplemented diets (*Albizia saman*, *Bridelia micrantha, Ficus sur,* or *Gmelina arborea*). These browse plants were shown in a concurrent but separate study to be reasonably nutritious (based on chemical composition and fiber constituents) and non-toxic (based on tannin, phytate, saponin, alkaloid and oxalate levels). Rumen liquids and solids for DNA extraction were collected via intubation from two animals in each group before and after dietary shift. Bacterial 16S rRNA gene regions V6-V8 were sequenced by 454 pyrosequencing. All communities were highly diverse and dominated by the phyla Firmicutes, Bacteroidetes, Tenericutes, Actinobacteria, and Proteobacteria. All communities shared members of the genera *Butryivibrio*, *Prevotella*, and *Ruminococcus.* Our analysis defined a core sets of bacteria shared by: all animals, forage-fed animals, and browse-fed animals. Community composition shifted dramatically in animals fed *A. saman* or *G. arborea*. The impact of tree-based browse on the ruminal bacterial community of Nigerian WAD sheep varies by browse species, likely due to differences in browse composition. Our study describes the first tropical small ruminant bacterial microbiome, and supports diet supplementation with specific tree-based browse for WAD sheep.

Introduction

There is an ever-increasing need to understand agricultural practices, especially for subsistence farmers in equatorial Africa. The West African Dwarf (WAD) sheep (*Ovis aries*, Djallonké) is an important meat animal reared on family farms [1] from Senegal to Botswana and number in the millions of animals per country [2]. WAD sheep are highly adaptable to a broad range of environments, can live on crop by-products without grain supplementation [3], are trypanotolerant [4], have rapid growth [5], and rich in genetic variation [6]. Many African farmers use tree-based browse as feed supplements, as these feed sources require no cultivation. Moreover, during the dry season, trees leaves and branches are potentially more nutritious than grasses [7,8] and with measurably higher crude protein content [9,10]. Previous work suggests that certain trees are acceptable feed substitutes with no detrimental effects on overall animal production [11,12,13]. However, some tree-based browse can result in feed refusal and loss of host nitrogen [13,14], possibly by adversely changing the ruminal microbial community through increased phenolic and tannin concentrations [15]. Thus it is important to investigate the effect of these diet alterations on animal digestion, particularly with respect to the host ruminal microbial community.

Ruminant digestion relies upon a ruminal microbial community composed of protozoa, fungi, bacteria, and archaea [16], with the bulk of cellulose hydrolysis thought to be performed by bacteria (for a recent review, see [17]). Recent diet-based studies characterizing rumen communities in cattle [18,19,20,21,22,23] and sheep [18,19,20] have highlighted the dynamic responses of ruminal microbes to changes in diet composition. To date, no study has been reported that utilizes a sequence-based approach to characterizing the ruminal bacterial microbiome for any tropical small ruminant.

As such, understanding the ruminal bacterial community and its response to diet supplementation is of particular interest in the tropics, where tree-based browse usage can decrease reliance on the production of grass-based forage. Moreover, recent work has suggested that for ruminants, a mixed diet including browse plants and forage can not only increase feed efficiency, but also promote higher biodiversity [21]. Here, we used Nigerian WAD sheep and 454 pyrosequencing to assess the impact of specific tree-based browse on tropical ruminal bacterial communities, with the hypothesis that tree-based browse diet supplementation would shift the bacterial communities to be distinct from grass-fed animals. We also investigated whether or not a core ruminal bacterial community is present in WAD sheep fed grass-based forage or tree-based browse. In this study, we fed WAD sheep a basal diet of the grass *Panicum maximum*, divided them into groups of three by weight, and kept one group on the basal diet while supplementing the diet of the other groups with one of the browse trees *Albizia saman*, *Bridelia micrantha, Ficus sur,* or *Gmelina arborea*. These tree species were chosen for their ubiquity and ease of growth in West Africa, current under-utilization as feed, and known lack of anti-nutritive toxicity [10] . We expect our results to be informative both for ruminant microbiologists and for future applications of tree-based browse in ruminant agriculture.

Materials and methods

Experimental feed materials and sources. Four indigenous or naturalized browse trees (*Albizia saman* (family *Fabaceae*), *Gmelina arborea* (*Verbenaceae*), *Ficus sur* (*Moraceae*), and *B. micrantha* (*Phyllanthaceae*)) were harvested from an arboretum, and the grass *Panicum maximum* from a cultivated pasture, established by the Department of Pasture and Range management of the Federal University of Agriculture, Abeokuta, Ogun State, Nigeria. The study

location (latitude: 7° N, longitude 3.5°) is in the savannah agro-ecological zone of southwest Nigeria in Abeokuta that receives an average annual rainfall of 1,037 mm [3]. A total of five diets were used in this study: Diet I (Control) - *P. maximum* (100 %); Diet II – *A. saman* (60 %) + *P. maximum* (40 %); Diet III – *G. arborea* (60 %) + *P. maximum* (40%); Diet IV – *F. sur* (60 %) + *P. maximum* (40 %); and Diet V – *B. micrantha* (60%) + *P. maximum* (40 %). Each diet was measured using dry matter, with total feed volume adjusted to prevent selective refusal in the mixed diets.

Feeding and Management of Animals. All animals were reared in the sheep unit at the Federal University of Agriculture, Nigeria, following institutional guidelines. Fifteen indigenous female West African Dwarf (WAD) sheep aged 5-7 months and weighing between 11.20 – 14.50 kg were grouped by weight into 5 cohorts of 3 animals. All animals were placed in disinfected individual pens with *ad libitum* water access, de-wormed with Albendazole® 2.5 % oral suspension at 1 mL/10 kg body weight and treated against ectoparasites with Cypermethrin® Pour-on at 1 mL/10 kg body weight. The weights of all animals were taken before the commencement of the experiment and weekly thereafter. Each cohort was fed exclusively *P. maximum* prior to the start of the experiment. Cohorts were then randomly assigned to a dietary treatment as described above, with all diets being offered in equal total quantities and refusals collected daily to track consumption of browse *vs*. *P. maximum* consumption. The entire feeding period lasted 84 days, with feed refusals collected and weighed daily after an initial 2-week adaptation period. A one-way ANOVA was performed on all animal metrics using the statistical software SAS, Version 6 (SAS Inst. Inc., Cary, NC).

Collection of rumen fluids and solids. Rumen contents were collected from two animals on each diet at the start (basal *P. maximum* forage diet) and completion (after the two-week

adjustment period after initiating diet supplementation with tree-based browse) of the experiment. All samples were collected prior to morning feeding. Each animal was assigned an arbitrary number (1 - 10) for sample identification purposes. An aliquot of 50 mL of rumen fluid containing solids was taken from the esophagus via suction tube. The rumen samples were filtered through four layers of cheesecloth to obtain both liquid and solid (fiber-adherent) portions of approximately 20 mL each. Each sample was preserved at -20 °C and shipped on dry ice to the University of Wisconsin-Madison, Madison, WI, USA for analysis of the ruminal microbiota. All samples were imported under USDA import permit #120106.

Template preparation and 454 pyrosequencing. Total genomic DNA was extracted separately from solid and liquid fractions using mechanical disruption with hot/cold phenol as previously described [22]. DNA quantification and integrity was measured using a Nanodrop (Thermo Scientific, Wilmington, DE) and by gel visualization (1 % agarose in TAE). PCR was carried out on each liquid and solid sample to amplify the V6-V8 variable region of the 16S rRNA gene using primers constructed from the universal 16S rRNA sequences 926F and 1392R coupled to the Roche 454 A or B Titanium sequencing adapters, respectively. Specifically, the forward primer used was 926F-5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGAAACTYAAAKGAATTGACGG-3' and the reverse primer included one of 20 barcodes, 5 bp in length, as indicated by XXXXX: 1392R-5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-XXXXX-ACGGGCGGTGTGTRC-3'. Each sample was amplified in 20 μL reactions containing 40 ng of DNA and 0.125 μM final concentration of each primer with the high-fidelity DNA polymerase Platinum Blue (Invitrogen Life Technologies, Grand Island, NY). The following PCR cycling conditions were used: initial denaturation of 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 the final extension at 68 °C for 10 min. Amplicon creation without secondary products was determined by gel electrophoresis (1 % agarose in TAE).

Total PCR products for liquid and solid DNA samples were combined to make a single equimolar pool for each animal. Each pool was cleaned twice with the Agencourt AMPure XP system (Beckman Coulter, Inc., San Diego, CA) to remove primers and short DNA fragments and then quantified using a Qubit® Fluorometer (Invitrogen,San Diego, CA). Amplicon quality in each pool was verified using an Agilent Bioanalyzer with the DNA 1000^{TM} chip (Agilent Technologies, Germany GmbH), and a final pool containing equimolar portions of all samples was made at 1×10^9 molecules/µL. An emPCR reaction was performed using an approximate ratio of 0.8 : 1 (amplicon : emPCR beads). Amplicon sequencing was performed following the manufacture's protocols (Roche Applied Science, Indianapolis, IN) for Titanium sequencing on a Roche 454 GS Junior Titanium sequencer using a Lib-L kit.

Sequence analysis. All sequences have been deposited with sample IDs and barcodes at the National Center for Biotechnological Information's Short Read Archive projects under accession SRP027328. Data analysis was performed using the bioinformatics program MOTHUR v.1.28.0 (Schloss *et al.* 2009). In brief, sequences were allowed to have a maximum of two differences in the primer and none in the barcode and de-noised using an implementation of the Amplicon Noise algorithm [23]. Sequences were trimmed to a minimum length of 250 bp, aligned against the SILVA 16S rRNA gene reference alignment database [24], and checked for putative chimeric sequences (*chimera.uchime*). All sequences were classified (*classify.seqs*) to operational taxonomic units (OTUs) at a 95 % identity level (*classify.otu*) using the Greengenes database [25] at a confidence level of at least 60 % with Cyanobacteria, Eukaryota, and Archaea lineages removed as our primers were not designed to amplify these groups. Sequence coverage was

determined using rarefaction and Good's coverage [26], with diversity measured using Simpson's diversity index [27]. The bacterial communities were analyzed using: AMOVA (iters = 1,000,000; Yue and Clayton Theta [28,29]), Good's coverage [26], principal coordinates analysis (PCA) [30], weighted UniFrac [31], Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering [32] based on the Morisita-Horn index [33], and as implemented in MOTHUR.

Results

All tree-based browses were consumed without negative host impact. Weight and feed consumption data for all study animals is reported in Table 1. None of the browse-supplemented diets (Diets II-V) had a negative impact on weight gain as compared to the control *P. maximum* diet (Diet I). In all cases, the addition of browse reduced the daily consumption of *P. maximum*. All dietary treatments resulted in positive body mass changes, with an average daily gain ranging from of 20.64 g (Diet I) to 41.72 g (Diet V). No animals showed signs of disease or ill-health during the course of the trial.

Sequencing the WAD sheep ruminal bacterial community. To assess each animal's ruminal bacterial community we performed a 454 pyrosequencing-based analysis of 16S rRNA genes on pooled (solids and liquids) rumen samples from 19 of our 20 samples. We used the variable regions V6-V8 in order to maximize sequence length and minimize the impact of any individual region on the final diversity and identity metrics. Because our primary interest was in the whole ruminal bacterial population and not on phase-specific shifts, and to maximize per-sample sequence coverage, we pooled the solid- and liquid-based PCR products. One sample (Animal 6 fed *F. sur*) did not survive transport and processing. From an initial set of 82,366 sequences, a

Table 1. Weight and feed intake metrics for all study animals (n = 3 per diet).

		Supplementation				
Average animal weight	Control Diet I (kg)	*A. saman* Diet II (kg)	*G. arborea* Diet III (kg)	*F. sur* Diet IV (kg)	*B. micrantha* Diet V (kg)	SEM
Initial	12.43[d]	14.27[b]	13.30[c]	14.50[a]	11.20[e]	0.32
Final	14.17[d]	16.43[b]	16.18[b]	17.50[a]	14.67[c]	0.33
Average daily feed intake*						
Tree-based browse	-	0.29[b]	0.30[b]	0.37[a]	0.35[a]	0.04
P. maximum	0.40[a]	0.19[c]	0.20[c]	0.24[b]	0.23[b]	0.02

[a, b, c, d, e] Significant within row ($P < 0.05$, ANOVA)

* Dry matter weight

total of 44,262 sequences were retained through all clean-up and filtering steps. Of these, 6,533 were unique, with an average length of 429 bp. Sequence distribution, Good's coverage, and Inverse Simpson's diversity index values are given in Table S1. An average of 2,330 ± 301 SD sequences per sample was obtained. Importantly, sufficient coverage for each sample was achieved, as measured by a Good's value of at least 92% for each sample (Table S1), a levelling of the associated rarefaction curves (Fig. S1), and the closeness of our final sequence counts to the theoretical maxima calculated from second order equations (all $R^2 > 0.97$) fitted to each rarefaction curve (Table S1). The inverse Simpson's index for all 10 animals (n=12 samples) on the control *P. maximum* diet ranged from 3.51 - 45.45, with an average of 19.8 ± 16.4 SD.

WAD sheep ruminal bacterial communities are small but rich in diversity. Among all animals, there were a total of 1,272 unique operational taxonomic units (OTUs) at 95 % sequence-similarity; 794 of these were classifiable to at least the family level with a minimum confidence of 60 %. The mean number of OTUs across all 10 animals fed *P. maximum* prior to browse supplementation was 260 ± 70 SD. These were dominated by sequences belonging to the phyla Firmicutes (57.0 %) and Bacteroidetes (17.9 %) (Fig. S2). The remaining sequences were distributed among the Tenericutes (7.6 %), unclassified Bacteria (8.8 %), Actinobacteria (2.7 %), Proteobacteria (1.7 %), and a number of low-abundance phyla. These *P. maximum*-fed animals were the only animals containing the phyla Fibrobacteres (0.3 %), Lentisphaera (0.3 %), and Fusobacteria (0.1 %). Other low-abundance phyla, such as TM7 and Chloroflexi, were found at similar levels in all diets, although there was individual variation by animal. The general pattern and order of relative OTU dominance was similar among all *P. maximum*-fed animals (Fig. S3A), but this pattern was less clear when examined by relative sequence abundance (Fig. S3B).

Table S1. Sequence distribution, coverage, and diversity of all WAD sheep rumen samples.

Diet	Tree-based browse	Animal	Sequence counts	95 % OTUs	Good's coverage	Calculated maximal sequences*	Inverse Simpson's diversity
I	None	9	2,191	326	0.93	2,714	45.45
	None	9	2,355	301	0.94	2,484	27.03
	None	10	2,811	361	0.94	3,468	43.48
	None	10	2,240	248	0.95	2,371	19.61
II	None	1	2,396	200	0.97	1,915	9.71
	A. saman	1	1,968	176	0.96	2,037	8.40
	None	2	2,433	236	0.95	2,363	8.55
	A. saman	2	2,237	161	0.97	2,426	5.29
III	None	7	1,860	117	0.97	2,750	3.51
	G. arborea	7	2,111	251	0.95	3,198	9.80
	None	8	1,867	324	0.92	1,871	40.00
	G. arborea	8	2,048	152	0.97	1,730	4.03
IV	None	5	2,946	231	0.97	2,535	7.81
	F. sur	5	2,370	208	0.97	2,560	8.00
	None	6	2,733	268	0.96	2,800	12.35
V	None	3	2,187	278	0.93	2,092	9.26
	B. micrantha	3	2,544	208	0.97	2,528	24.39
	None	4	2,456	247	0.97	2,382	17.86
	B. micrantha	4	2,509	279	0.96	2,093	13.51

* Number of total sequences needed for the maximum recovery of unique sequences calculated from a second-order polynomial best-fit equation

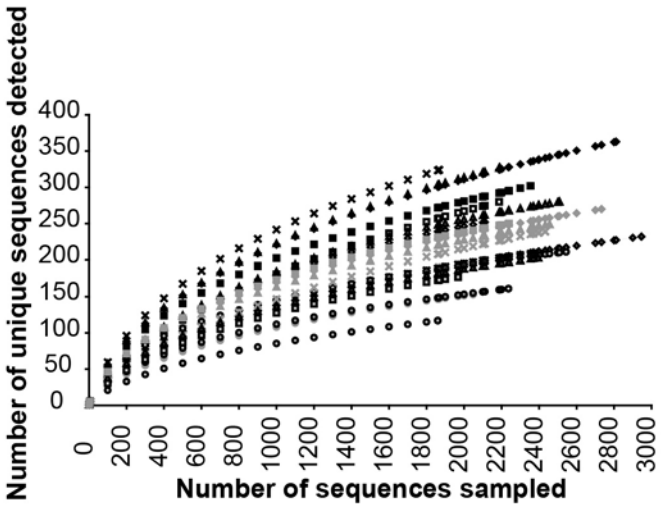

Figure S1. Rarefaction curves for all samples at the 95 % OTU level.

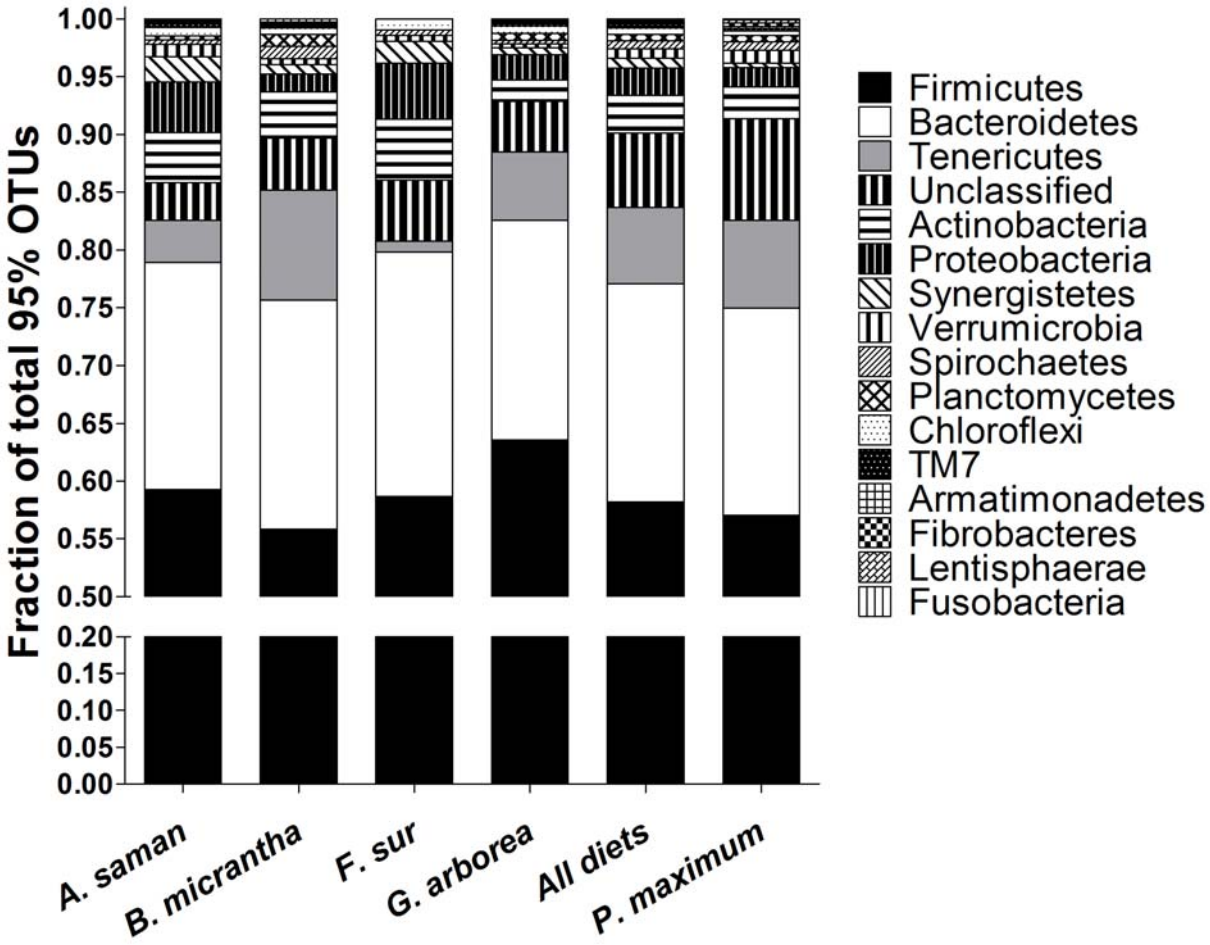

Figure S2. Comparison of phylum-level distribution of 95% OTUs by relative abundance of sequences. The *P. maximum*-fed group includes both early and late time points for animals 9 and 10. *P. maximum*-fed (Diet I) N = 12, *A. saman*-fed (Diet II) N = 2, *G. arborea* (Diet III) N = 2, *F. sur*-fed (Diet IV) N = 1, *B. micrantha*-fed (Diet V) N = 2, all diets N = 19.

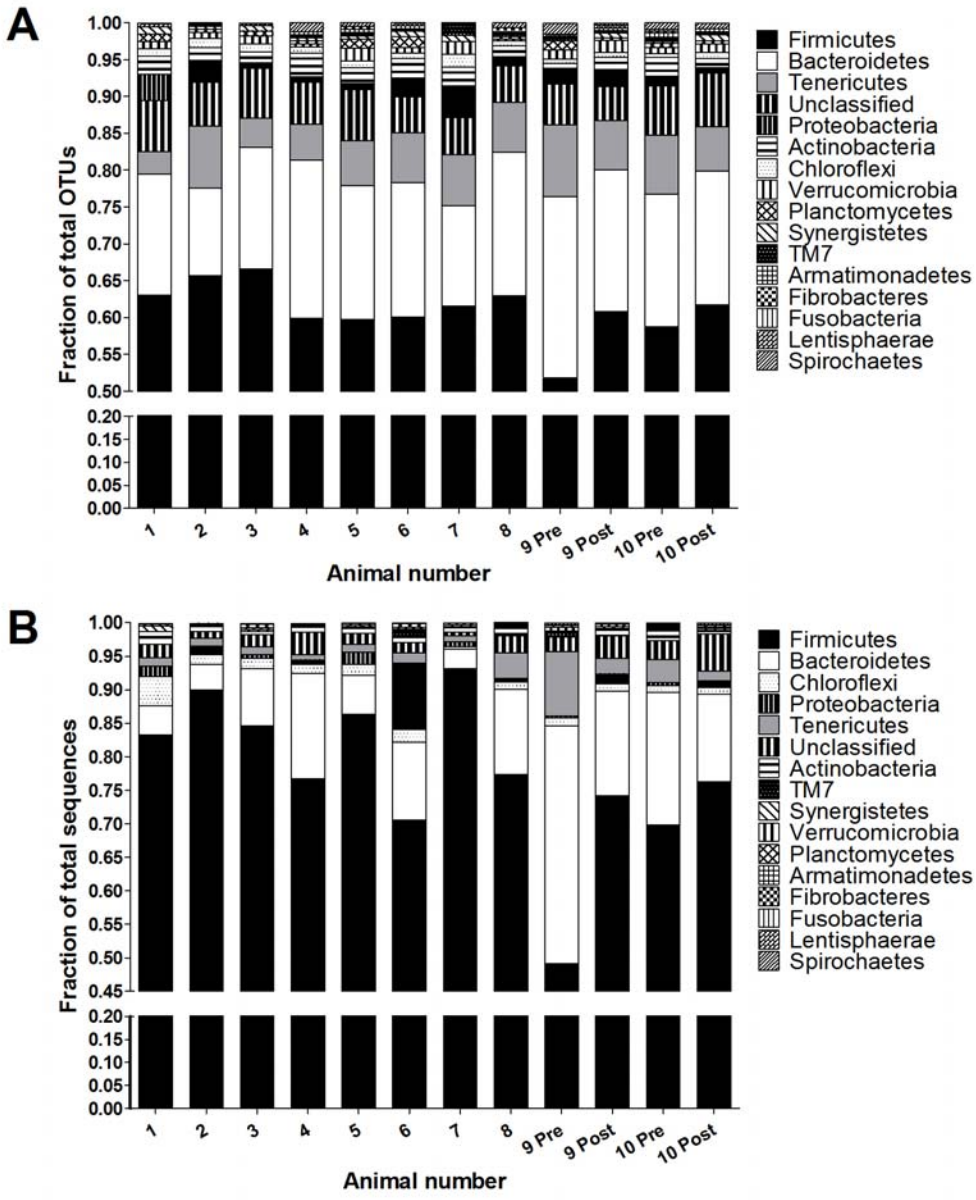

Figure S3. Comparison of phylum-level distribution of 95 % OTUs by relative abundance of OTUs and sequences for all *P. maximum*-fed animals. Animals are identified by number, with control animals 9 and 10 having two entries due to being sampled both when all animals were fed *P. maximum* ("Pre") and after all other animals (1-8) underwent diet supplementation ("Post"). (A) Phylum-level distribution of 95 % OTUs based upon relative abundance of OTUs; (B) Phylum-level distribution of 95 % OTUs based upon relative abundance of sequences within OTUs.

Both analysis methods indicated that the dominant phyla are the Firmicutes and Bacteroidetes for all *P. maximum*-fed animals.

When analyzed at 95 % OTUs (approximately genus-level) the individual ruminal bacterial communities diverged between both treatments and animal pairs. There was a high degree of variation in the total community composition among all Diet I samples (Fig. 1A & 1B), with each animal's total bacterial community significantly different from all others as determined by a weighted UNIFRAC ($P < 0.001$). When the total community composition was analyzed by UPGMA (Fig. 1A) or PCA (Fig. 1B), there was a general pattern of *P. maximum*-fed animals separating from the tree browse-fed animals, but this included a high degree of mixing. Although the large total loading values of our PCA plot (73.18 % of total variance) indicates that the community pattern could be explained by two major variables, there is a lack of clustering by our chosen variable of diet treatment.

Feed supplementation can alter the individual ruminal communities. To determine the impact of tree-based browse on the total ruminal community composition in each animal we performed a vector analysis for each animal using our PCA coordinates (pre- and post-diet supplementation) (Fig. 1C) and an AMOVA for each diet. By PCA vector comparison, each animal pair's vector magnitudes showed two major groups of communities: those with magnitudes less than 0.103 (Diets I, IV, and Diet V) and those with magnitudes greater than 0.270 (Diets II and III). Given that larger magnitudes correspond to a greater degree of total community composition change, our results indicate that Diets II and III resulted in dramatic shifts within the WAD sheep rumen community, regardless of the community's starting structure (Fig. 1B). Diets IV and V resulted in a degree of community shift, as measured by vector

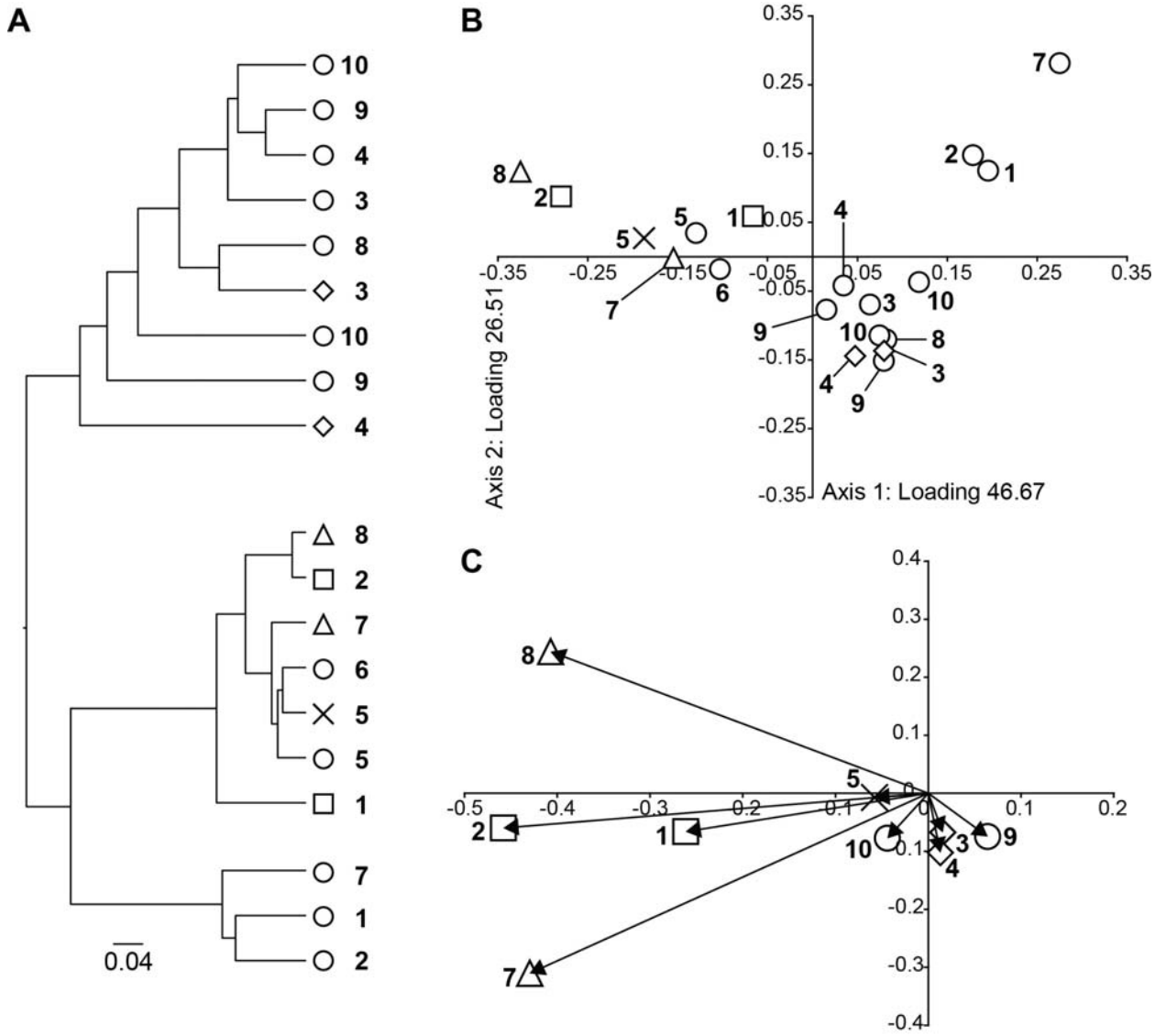

Figure 1. Comparison of the West African Dwarf sheep ruminal community composition by UPGMA, PCA, and vector analysis. (A) UPGMA tree, and (B) PCA plot showing total community composition relationships based on 95 % similarity (0.05) OTUs. (C) The degree of change for each sample pair (from before to after diet supplementation) was calculated as a vector from their respective PCA coordinates. Individual animals are identified by number (1-10). Diets: (○) I control, (□) II, (Δ) III, (×) IV, (◊) V.

analysis, comparable to the animal pair maintained on the control Diet I. By AMOVA only Diet V was not significantly different from the control animals ($P = 0.333$ for Diet V and $P < 1\times10^{-6}$ for all other diets). The community composition changed significantly ($P < 1\times10^{-6}$, AMOVA) for each animal over the course of the experiment except for the single *F. sur*-supplemented animal.

To measure specific changes occurring within these communities, we analyzed the distribution of OTUs among all diets. In order to reduce inter-animal variation we required each OTU to be present in half or more of the samples for Diet I. For Diets II-V we required the OTU be present in all animals on that diet. Unclassified OTUs were not considered in our analysis. All OTUs were pooled at the genus level and the relative abundance of OTUs within each genus was calculated relative to the total number of OTUs in each diet (Table 2). By using an OTU-based abundance metric we were able to approximate a measure of the diversity present within each genus. In all diets, the highest percentage of OTUs was within *Prevotella* (16.67 - 33.33 %). Decreases in OTU diversity, relative to Diet I, was seen for *Eubacterium* (5.4-fold) and *Ruminococcus* (2.9-fold) on Diet IV, while *Ruminococcus* and *Coprococcus* decreased 2.3-fold in Diet II. Changes greater than 2-fold, relative to Diet I, included *Oribacterium* (4.4-fold) in Diet III and *Shuttleworthia* (4.1-fold) and *Oscillospira* (3.1-fold) in Diet V.

In order to determine changes in relative abundance of highly represented genera we compared the ten most abundant genera in each diet (Fig. S4). In many cases both animals on each diet showed either an increase or decrease relative to Diet I, but with only one of the two animals was outside of the expected level of variation (calculated as the standard deviation of the mean in all 12 Diet I control samples). In the following cases both animals on a given diet had the same trend and were outside of the expected variation: *Carnobacterium* increased in Diets II,

Table 2. Percent relative sequence abundance distribution of genus-level OTUs.

	Control	Supplementation (n=2 samples per diet)		
	Diet I n=12 samples	*A. saman* Diet II	*G. arborea* Diet III	*B. micrantha* Diet V
Shared OTUs*	92	35	42	45
Prevotella	27.17	17.14	16.67	33.33
Eubacterium	11.96	14.29	11.90	2.22
Butyrivibrio	7.61	11.43	7.14	6.67
Clostridium	6.52	-	4.76	6.67
Coprococcus	6.52	2.86	7.14	4.44
Ruminococcus	6.52	2.86	9.52	2.22
Acetivibrio	4.35	-	-	6.67
Selenomonas	3.26	2.86	-	2.22
Oscillospira	2.17	-	2.38	6.67
Lactococcus	2.17	2.86	2.38	-
SHD-231	2.17	-	4.76	4.44
Bulleidia	2.17	5.71	-	4.44
Desulfovibrio	1.09	2.86	2.38	2.22
p-75-a5	1.09	-	2.38	2.22
Shuttleworthia	1.09	2.86	2.38	4.44
Carnobacterium	1.09	2.86	2.38	-
Enterococcus	1.09	2.86	2.38	-
Psychrobacter	1.09	2.86	2.38	-
Streptococcus	1.09	2.86	2.38	-
Pseudobutyrivibrio	1.09	-	2.38	-
Staphylococcus	1.09	-	2.38	-
Oribacterium	1.09	2.86	4.76	2.22
Dehalobacterium	1.09	-	-	2.22
TG5	1.09	-	-	2.22
Treponema	1.09	-	-	2.22
Atopobium	1.09	2.86	-	-
Aerococcus	1.09	-	-	-
L7A_E11	1.09	-	-	-
Arthrobacter	-	-	2.38	-
Paludibacter	-	-	2.38	-
Planomicrobium	-	-	2.38	-
RFN20	-	8.57	-	2.22
Pseudomonas	-	2.86	-	-
Succinivibrio	-	2.86	-	-
Synergistes	-	2.86	-	-

* Shared by at least 6 samples for *P. maximum*-fed animals and by both animals for other diets (Diet IV, *F. sur*, is not included due to being a singlet).

III, and IV; *Psychrobacter* increased in Diet IV; *Staphyloccocus* increased in Diet II; and SHD-231 (family Anaerolinaceae) decreased in Diet V.

WAD sheep have a core microbiome. We then identified the set of core OTUs present in WAD sheep across all animals on all diets (Diets I-V); those fed tree-based browse (Diets II-IV); and those fed only *P. maximum* grass forage (Diet I and initial samples from all animals). Many of the OTUs share genus- or family-level taxonomy, and included both classical ruminal genera such as *Ruminococcus* and unclassified genera. As shown in Table 3, there were 11 OTUs present in all samples, the most abundant of which belonged to an unclassified member of the Catabacteriaceae. Other members of the shared core microbiome include OTUs in the Chloroflexi, Clostridiales, Ruminococcaceae, *Butyrivibrio*, *Prevotella*, and *Ruminococcus*. The core set of OTUs found in all animals on all diets were not significantly different in relative abundance when split by forage *vs*. browse diets (1-way ANOVA, $P > 0.05$).
OTUs present only in all browse-supplemented animals included six OTUs, of which the most abundant was an unclassified member of the Lachnospiraceae. The other members of this browse-fed core OTU set belong to the Bacteroidetes, Veillonellaceae, *Prevotella*, and *Desulfovibrio*. The animals fed only *P. maximum* had 11 OTUs present in all samples that did not appear in any animals fed Diets II-V, with the highest relative sequence abundance OTU being a member of *Staphylococcus* at 2.24 %. The remaining OTUs included members of the Chloroflexi, Tenericutes, Catabacteriaceae, Clostridiales, Lachnospiraceae, *Atopodium*, and *Selenomonas*.

Classical ruminal bacteria are also present in the WAD sheep rumen. Given the known functional importance of classical ruminal bacteria in other ruminants, we determined if there existed a correlation between these bacteria and community shifts on the various tree-based

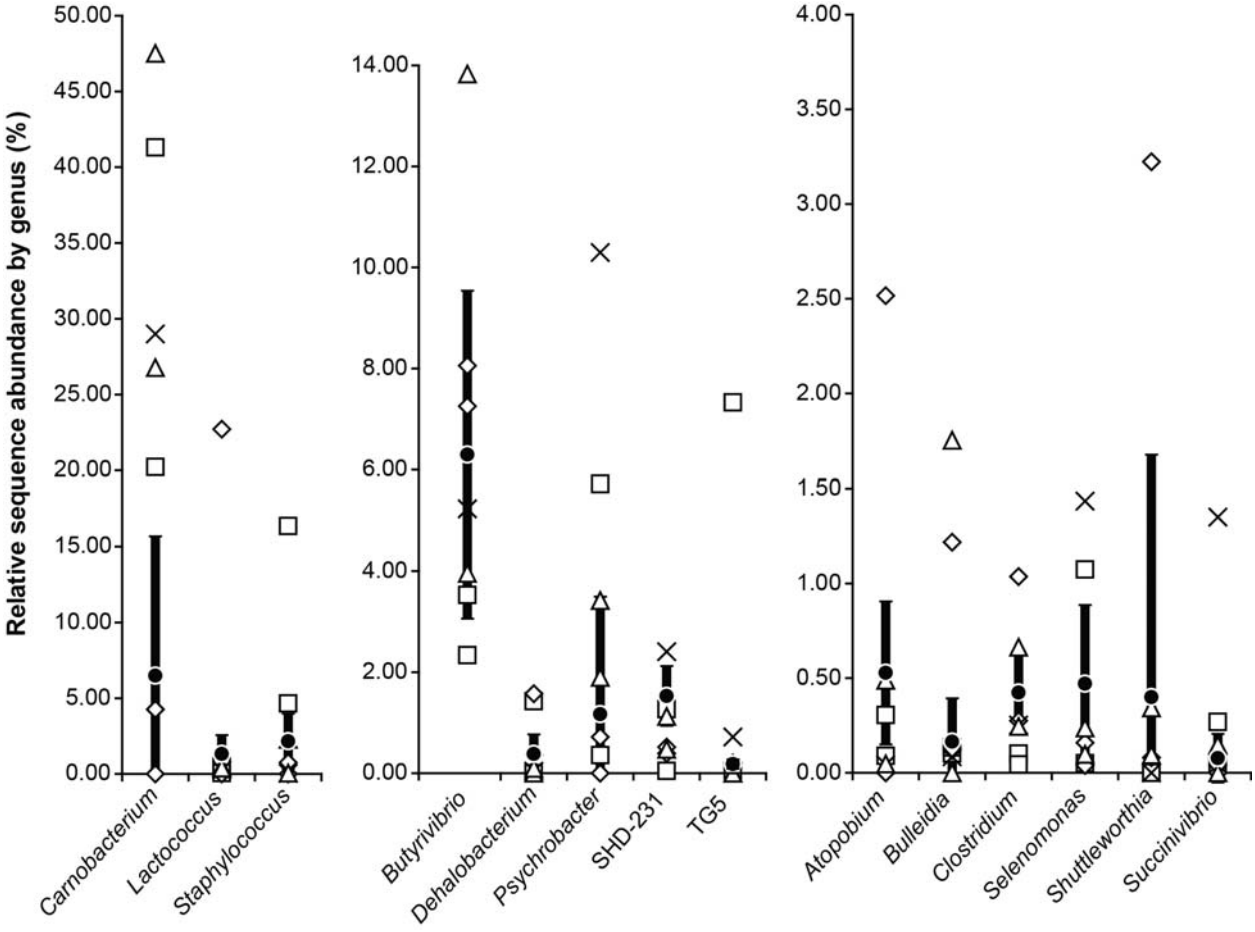

Figure S4. Differences in relative sequence abundance for the most abundant genera based on diet. The relative sequence abundance for classifiable genera with > 1 % abundance in at least one animal is given for each animal on a tree-based browse diet. The mean abundance with standard deviation (black bars) was calculated from the pooled samples taken from all *P. maximum*-fed animals (n=12, Diet I). Diets: (●) I [control], (□) II, (Δ) III, (×) IV, (◊) V.

Table 3. Percent relative sequence abundance of core ruminal bacterial communities found in West African Dwarf sheep.

Control Diet I (n=12)*	*A. saman* Diet II (n=2)	*G. arborea* Diet III (n=2)	*F. sur* Diet IV (n=1)	*B. micrantha* Diet V (n=2)	OTU (confidence)†
Present in all animals, times, and diets (Diets I-V)					
14.61	14.03	13.34	16.58	10.55	Firmicutes: Catabacteriaceae (96)
4.77	2.24	6.40	2.66	6.02	Firmicutes: *Butyrivibrio* (100)
2.85	2.12	1.83	1.31	6.23	Firmicutes: Ruminococcaceae (100)
2.08	0.81	1.80	1.31	4.51	Firmicutes: Clostridiales (94)
1.08	0.62	0.31	0.51	1.70	Firmicutes: Clostridiales (100)
0.93	0.38	0.17	2.03	0.22	Chloroflexi: SHD-231 (100)
0.64	0.74	1.27	0.13	0.79	Firmicutes: Clostridiales (92)
0.80	0.19	0.26	1.18	0.46	Bacteroidetes: *Prevotella* (100)
0.50	0.33	0.79	0.30	0.77	Firmicutes: *Butyrivibrio* (100)
0.57	0.26	0.63	0.13	0.38	Firmicutes: Clostridiales (100)
0.58	0.48	0.22	0.34	0.30	Firmicutes: *Ruminococcus* (100)
Present only in all browse-supplemented animals (Diets II-V)					
-	0.43	0.48	0.34	0.77	Firmicutes: Lachnospiraceae (100)
-	0.45	0.67	0.59	0.12	Bacteroidetes (100)
-	0.71	0.14	0.13	0.40	Firmicutes: Veillonellaceae (95)
-	0.05	0.07	0.13	0.04	Bacteroidetes: *Prevotella* (100)
-	0.21	0.05	0.04	0.14	Proteobacteria: *Desulfovibrio* (100)
-	0.10	0.12	0.08	0.12	Firmicutes: Lachnospiraceae (100)
Present only in *P. maximum*-fed animals both pre- and post-supplementation					
2.24	-	-	-	-	Firmicutes: *Staphylococcus* (100)
1.72	-	-	-	-	Firmicutes: Lachnospiraceae (100)
1.73	-	-	-	-	Firmicutes: Catabacteriaceae (100)
0.62	-	-	-	-	Chloroflexi: SHD-231 (100)
0.50	-	-	-	-	Actinobacteria: *Atopodium* (73)
0.46	-	-	-	-	Firmicutes: Clostridiales Family XIII Incertae Sedis (100)
0.26	-	-	-	-	Tenericutes: p-75-a5 (100)
0.20	-	-	-	-	Firmicutes: Lachnospiraceae (88)
0.17	-	-	-	-	Firmicutes: Lachnospiraceae (100)
0.16	-	-	-	-	Firmicutes: Clostridiales (100)
0.19	-	-	-	-	Firmicutes: *Selenomonas* (99)

* Includes 10 samples from all animals prior to diet supplementation and two samples from the control-diet animals post diet supplementation.

† Taxonomic identifications for each OTU are given at the phylum level followed by the most specific classification that could be defined, with the confidence for each classification given in parentheses.

browse supplemented diets. Specifically, we examined sequence abundances for Diets II - V, relative to Diet I, that were classified as belonging to the following important ruminal genera: *Bacteroides* [34], *Butyrivibrio* [34], *Fibrobacter* [35], *Megasphaera* [36], *Prevotella* [35], *Ruminococcus* [37], *Selenomonas* [34], and *Streptococcus* [38]. All results are given in Table S2, with fold changes considered significant when more than 2-fold. The genera *Fibrobacter*, *Lachnospira*, and *Megasphaera* were in extremely low abundance in all diets, and were not detectable in most samples. In Diet II *Bacteroides*, *Butyrivibrio*, *Prevotella*, and *Streptococcus* decreased, while *Succinivibrio* increased. In Diet III *Bacteroides*, *Prevotella*, and *Selenomonas* were decreased. In Diet IV *Lachnospira*, *Selenomonas*, *Streptococcus*, and *Succinivibrio* all increased, with *Succinivibrio* being nearly 20-fold higher. In Diet V, which was the diet that by other metrics resulted in a ruminal community most closely resembling that of the control diet, there were decreases in *Bacteroidetes*, *Selenomonas* and *Succinivibrio*.

Discussion

The ability to shift livestock feeding practices to local browse plants has the potential to increase the availability of arable land for other agricultural uses such as the cultivation of cash crops. Importantly, such a shift should not negatively impact livestock health or production. Of the diets compared in this study, only the control diet of *P. maximum* is a grass (guineagrass, Diet I), while all browse plants are trees indigenous or naturalized to tropical Africa that grow without deliberate cultivation on sub-optimal land. We found that all tree-based browse diets tested were accepted as feed by WAD sheep, and none negatively impacted weight gain or health over the course of our study.

Our WAD sheep rumen community analysis revealed that the total bacterial community for each animal was small, relative to other ruminants like cattle [39,40,41], but remained highly diverse. This may reflect a tightly-knit ruminal community enabling WAD sheep to flourish on sub-optimal feeds, or rumen sample degradation during shipment. In particular, we found that the WAD sheep rumen is dominated by the Firmicutes, Bacteroidetes and Tenericutes, with the minor presence of Actinobacteria and Proteobacteria (Fig. S1). Previous work in sheep using clone libraries found similar trends for the Firmicutes and Bacteroidetes [25,26,48], and a recent multi-species (cows, sheep, and red deer) pyrosequencing-based study found that among all samples the phyla with the most abundant families were the Firmicutes, Bacteroidetes, and Fibrobacteres [42]. In our data set the high abundance of sequences and OTUs in the Tenericutes suggests that this phylum may play a unique role in WAD sheep.

Our analysis also revealed 11 OTUs shared across all animals and diets, of which many are well-known ruminal bacteria including Clostridiales, Ruminococcaceae, *Butyrivibrio*, and *Prevotella* (Table 3). Bacteria in the hemicellulolytic genera *Prevotella* and *Butyrivibrio* accounted for at least 25 % of the sequences recovered from all of our samples, whereas bacteria in the cellulolytic genera *Ruminococcus* and *Fibrobacter* contributed to at most 9.5 % of the sequences in any given diet (Table 2). Comparisons between grass and browse-supplemented diets showed only modest increases and decreases in the abundances of these bacteria (Table S2), suggesting that their populations were stable throughout the experiment. We also found a member of the Catabacteriaceae and the Chloroflexi bacterium SHD-231 conserved across all diets; members of these groups have likewise been found in other ruminants [43,44,45]. Given that all of these OTUs persisted after diet shifts, it is likely that these bacteria play important

roles within the WAD sheep ruminal community, possibly by participating in the fermentation of plant polysaccharides and production of volatile fatty acids.

Our data also show clear shifts in the bacterial community composition. It is unlikely that these shifts are due to changes in the ruminal community due to maturation of the animals over the course of the feeding trial, as the degree of change for the control diet animals (and *B. micrantha*) was extremely small (Fig. 1). Differences in ruminal flora due to browse supplementation may instead be due to compositional differences between *P. maximum* and the browse plants used in our study. These feeds are similar (0.6-1.5-fold relative to *P. maximum*) in levels of cellulose, lignin, neutral detergent fiber, and dry matter [9,10]. *B. micrantha* was the only feed with an elevated level of acid detergent fiber (2.0-fold) [9,10]. All of the tree-based browses were higher in crude protein (1.8 to 2.5-fold) and lower in hemicelluloses (2.0 to 4.9-fold) except for *B. micrantha*, which has a hemicellulose composition similar to *P. maximum* (1.1-fold) [9,10] and which did not significantly change the ruminal bacterial community as compared to *P. maximum* during our study. It is also possible that specific anti-nutritional compounds (such as alkaloids) could have an impact on the ruminal bacterial community. However, we did not see this effect, as the diet most like the *P. maximum* control, *B. micrantha*, has elevated levels of tannin, phytate, alkaloid, and oxalate [9,10]. Of the seven OTUs shared between *P. maximum* and *B. micrantha*-fed animals five were in the Clostridiales. The shared Clostridiales OTUs included *Eubacterium*, two Lachnospiraceae, two Clostridiales, and one Catabacteriaceae; many bacteria in these groups are known to be involved in plant matter degradation, including hemicellulose fermentation specialists. It is probable that feed quality, digestibility, and freshness all play significant roles in impacting the ruminal community, and the impact of these factors remains to be investigated.

Our whole-community composition analysis revealed that browse-supplemented diets resulted in increased OTU diversity, although there was a high degree of inter-animal variation per diet (Fig. 1). We suspect that this high degree of variation is due to the complex genetic diversity of the WAD sheep, and the free-range flock management strategy used outside of the feeding trial. Of the diets tested, we found that Diets IV and V were comparable to the control diet (Fig. 1) with Diet V having a total community composition indistinguishable from the control diet. This finding suggests that, instead of shifting the ruminal community to better digest the browse plants, these diet supplements were best suited to take advantage of the bacteria already present in the WAD sheep rumen.

In conclusion, we have defined a core set of bacterial OTUs for the WAD sheep based on the variable regions V6-V8 of the 16S rRNA gene, and further characterized the major shifts in the ruminal populations due to diet supplementation with multiple tree-based browses. We acknowledge that the number of animals used for individual diets other than *P. maximum* is relatively small, as we were restricted by available resources related to sampling, storing, and shipping specimens between Nigeria and the USA. One further area of investigation not addressed in our research, but of potential future interest, is that methanogens in sheep can be suppressed by increasing dietary protein [46]. The browse used in our study, with nearly twice the protein of *P. maximum* [9,10], may have a similar impact on WAD sheep. Similarly, one of the classifiable OTUs conserved among only browse-supplemented diets, *Desulfovibrio*, are known members of the sheep rumen community that reduce sulphate to sulphide [47] and can act as competitors for methanogens [48]; conversely they may also encourage the growth of methanogens through the production of hydrogen in low-sulfate conditions [49], such as are found in the rumen. In order to fully understand the impact of diet upon the rumen community,

future studies should include methanogens, protists, and fungi, in addition to bacteria. The analysis presented here can be used as a framework for advancing our knowledge of the general ruminal microbial community composition in tropical small ruminants, in addition to providing a framework useful for the management and understanding of WAD sheep agriculture.

Acknowledgements

This work was supported by the University of Wisconsin-Madison College of Agriculture and Life Sciences and a USDA NIFA Fellowship Grant (2012-01193) to K. Jewell. All authors wish to thank the members of the sheep unit at the Federal University of Agriculture for their support during the feeding trial. We also wish to thank all members of the Suen Lab for their support, insightful discussions, and careful reading of the manuscript.

References

1. Adesehinwa A, Okunola J, Adewumi M (2004) Socio-economic characteristics of ruminant livestock farmers and their production constraints in some parts of South-western Nigeria. Livestock Research for Rural Development 16: Article #61.
2. Shaw A, Hendrickx G, Gilbert M, Mattioli R, Codjia V, et al. (2006) Mapping the benefits: a new decision tool for tsetse and trypanosomiasis interventions. International Development, Animal Health Programme. Rome, Italy: Food and Agriculture Organization of the United Nations.
3. Anele U, Arigbede O, Südekum K-H, Ike K, Oni A, et al. (2010) Effects of processed cowpea (*Vigna unguiculata* L. Walp) haulms as a feed supplement on voluntary intake, utilization and blood profile of West African dwarf sheep fed a basal diet of *Pennisetum purpureum* in the dry season. Animal Feed Science and Technology 159: 10-17.
4. Geerts S, Osaer S, Goossens B, Faye D (2009) Trypanotolerance in small ruminants of sub-Saharan Africa. Trends in Parasitology 25: 132-138.
5. Sowande O, Sobola O (2007) Body measurement of West African dwarf sheep as parameter for estimation of live weight. Tropical Animal Health and Production 40: 433-439.
6. Akinyemi M, Salako A (2010) Hemoglobin polymorphism and morphometrcal correlates in the West African Dwarf sheep of Nigeria. International Journal of Morphology 28: 205-208.
7. Aletor V, Omodara O (1994) Studies on some leguminous browse plants with particular reference to their proximate, mineral and some endogenous anti-nutritional constituents. Animal Feed Science and Technology 46: 343-348.
8. D'Mello J, Fraser K (1981) The composition of leaf meal from *Leucaena leucocephala*. Tropical Science 23: 75-78.
9. Odedire J, Babayemi O (2008) Comparative studies on the yield and chemical composition of Panicum maximum and Andropogon gayanus as influenced by Tephrosia candida and Leucaena leucocephala. Livestock Research for Rural Development 20: Retrieved April 25, 2013, from http://www.lrrd.org/lrrd2020/2012/oded20027.htm
10. Omoniyi L, Isah O, Olorunsola R, Osofowara O, Akinbode R, et al. (2013) Nutrient and anti-nutritional constituents of *Penisetum purpureum* and four indigenous tree legume of South-Western Nigeria: a potential ruminant feed. Nigerian Journal of Animal Production 40: 152-160.
11. Bamikole M, Ikhatua U, Arigbede O, Babayemi O, Etela I (2004) An evaluation of the acceptibility as forage of some nutritive and antinutritive components and of the dry matter degradation profiles of five species of *Ficus*. Tropical Animal Health and Production 36: 157-167.
12. Ouédraogo-Koné S, Kaboré-Zoungrana C, Ledin I (2008) Intake and digestibility in sheep and chemical composition during different seasons of some West African browse species. Tropical Animal Health and Production 40: 155-164.
13. Osakwe II, Drochner W (2006) Nutritive value of *Morinda lucida* and its fermentation parameters in West African dwarf (WAD) sheep when fed as supplement to grass hay. Small Ruminant Research 64: 107-115.
14. Reed JD, Soller H, Woodward A (1990) Fodder tree and straw diets for sheep: intake, growth, digestibility and the effects of phenolics on nitrogen utilisation. Animal Feed Science and Technology 30: 39-50.

15. Waterman P, Mbi C, McKey D, Gartlan JS (1980) African rainforest vegetation and rumen microbes: Phenolic compounds and nutrients as correlates of digestibility. Oecologia 47: 22-33.
16. Dehority BA (2003) Rumen Microbiology. Thrumpton, Nottingham: Nottingham University Press.
17. Wilson DB (2011) Microbial diversity of cellulose hydrolysis. Current Opinion in Microbiology 14: 259-263.
18. Saro C, Ranilla MJ, Carro MD (2012) Postprandial changes of fiber-degrading microbes in the rumen of sheep fed diets varying in type of forage as monitored by real-time PCR and automated ribosomal intergenic spacer analysis. Journal of Animal Science 90: 4487-4494.
19. Stiverson J, Morrison M, Yu Z (2011) Populations of select cultured and uncultured bacteria in the rumen of sheep and the effects of diets and ruminal fractions. International Journal of Microbiology 2011: doi:10.1155/2011/750613.
20. Perumbakkam S, Mitchell EA, Morrie Craig A (2011) Changes to the rumen bacterial population of sheep with the addition of 2,4,6-trinitrotoluene to their diet. Antonie van Leeuwenhoek Journal of Microbiology 99: 231-240.
21. Broom DM, Galindo FA, Murgueitio E (2013) Sustainable, efficient livestock production with high biodiversity and good welfare for animals. Proceedings of the Royal Society B: Biological Sciences 280: 10.1098/rspb.2013.2025.
22. Stevenson DM, Weimer PJ (2007) Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Applied Microbiology and Biotechnology 75: 165-174.
23. Quince C, Lanzen A, Davenport R, Turnbaugh P (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12: 38.
24. Pruesse E, Quast C, Knittle K, Fuches B, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35: 7188-7196.
25. DeSantis T, Hugenholtz P, Larsen N, Rojas M, Brodie E, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench campatible with ARB. Applied and Environmental Microbiology 72: 5069-5072.
26. Good I (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40: 237-264.
27. Simpson E (1949) Measurement of diversity. Nature 163: 688.
28. Yue J, Clayton M (2005) A similarity measure based on species proportions. Communications in Statistics - Theory and Methods 34: 2123-2131.
29. Exchoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479-491.
30. Gower J (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 325-338.
31. Lozupone C, Hamady M, Knight JR (2006) UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 7: 371.
32. Sokal R, Michener C (1958) A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38: 1409-1438.

33. Horn H (1966) Measurement of "overlap" in comparative ecological studies. The American Naturalist 100: 419-424.
34. Wallace R, Brammall M (1985) The role of different species of bacteria in the hydrolysis of protein in the rumen. Journal of General Microbiology 131: 821-832.
35. Osborne JM, Dehority BA (1989) Synergism in degradation and utilization of intact forage cellulose, hemicellulose, and pectin by three pure cultures of ruminal bacteria. Applied and Environmental Microbiology 55: 2247-2250.
36. Counotte GHM, Prins RA, Janssen RHAM, deBie MJA (1981) Role of *Megasphaera elsdenii* in the fermentation of dl-[2-^{13}C]lactate in the rumen of dairy cattle. Applied and Environmental Microbiology 42: 649-655.
37. Sijpesteijn AK (1951) On *Ruminococcus flavefaciens*, a cellulose-decomposing bacterium from the rumen of sheep and cattle. Journal of General Microbiology 5: 869-879.
38. Hudson JA, Cai Y, Corner RJ, Morvan B, Joblin KN (2000) Identification and enumeration of oleic acid and linoleic acid hydrating bacteria in the rumen of sheep and cows. Journal of Applied Microbiology 88: 286-292.
39. de Oliveira MNV, Jewell KA, Freitas FS, Benjamin LA, Tótola MR, et al. (2013) Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. Veterinary Microbiology 164: 307-314.
40. Li RW, Connor EE, Li C, Baldwin VIRL, Sparks ME (2012) Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. Environmental Microbiology 14: 129-139.
41. Shanks OC, Kelty CA, Archibeque S, Jenkins M, Newton RJ, et al. (2011) Community structures of fecal bacteria in cattle from different animal feeding operations. Applied and Environmental Microbiology 77: 2992-3001.
42. Kittelmann S, Seedorf H, Walters W, Clemente JC, Knight R, et al. (2013) Simultaneous amplicon sequencing to explore co-occurence patterns of bacterial, archaeal, and eukaryotic microorganisms in rumen microbial communities. PLOS ONE 8: e47879.
43. Li RW, Wu S, Li W, Huang Y, Gasbarre LC (2011) Metagenome plasticity of the bovine abomasal microbiota in immune animals in response to *Ostertagia ostertagi* infection. PLOS ONE 6: e24417.
44. Samsudin AA, Evans PN, Wright A-DG, Al Jassim R (2011) Molecular diversity of the foregut bacteria community in the dromedary camel (*Camelus dromedarius*). Environmental Microbiology 13: 3024-3035.
45. Kong Y, Teather R, Forster R (2010) Composition, spatial distribution, and diversity of the bacterial communities in the rumen of cows fed different forages. FEMS Microbiology Ecology 74: 612-622.
46. McAllister T, Newbold C (2008) Redirecting rumen fermentation to reduce methanogenesis. Australian Journal of Experimental Agriculture 48: 7-13.
47. Howard BH, Hungate RE (1976) *Desulfovibrio* of the sheep rumen. Applied and Environmental Microbiology 32: 598-602.
48. Ellis J, Dijkstra J, Kebreab E, Bannink A, Odongo N, et al. (2008) Aspects of rumen microbiology central to mechanistic modeling of methane production in cattle. Journal of Agricultural Science 146: 213-233.
49. Bryant MP, Campbell LL, Reddy CA, Crabill MR (1977) Growth of *Desulfovibrio* in lactate or ethanol media low in sulfate in association with H_2-utilizing methanogenic bacteria. Applied and Environmental Microbiology 33: 1162-1169.

CHAPTER FIVE

Summative discussion

In this dissertation I present work detailing the ruminal bacterial community within the framework of three major questions. The answers to these questions, dealing with the shared ruminal community, the correlation of community membership and structure with host efficiency, and the impact of diet on the community, improve both our understanding of ruminant microbiology and the interplay between the ruminant and its ruminal bacteria. In addition my results can act as the basis for future studies aimed at creating predictive models for the ruminal community and cow production efficiency.

Similar methods were used throughout the studies described in this dissertation for sample processing, targeting of the 16S rRNA V6-V8 region, sequencing by 454 pyrosequencing on a GS Junior machine, and sequence processing with the program mothur. However the differences in amplicon creation and preparation, specifically the DNA polymerase used (see Appendix 2) and gel extraction or bead-based size exclusion of the amplicon pools [1], mean that comparisons of low abundance phyla (< 0.3 %), and single-sequence OTUs, will not be performed. With appropriate care, I herein discuss patterns discernable across all three studies.

The shared ruminal bacterial community is a minor component of the entire, and highly variable, total community

Broadly, the dominant ruminal bacterial phyla are alternately reported as Firmicutes [2,3,4,5,6,7,8] or Bacteroidetes [7,9,10,11,12], probably due to differences between the host study species and feeding conditions. The ranking of the remaining phyla is not consistent between studies, but the rumen can include a high relative abundance (> 1 %) of Fusobacteria [6], Proteobacteria [2,5,8], Spircochetes [8], Synergistetes [2], [2,5], TM7 [6], and Tenericutes [6,8]. In the work of this dissertation the most abundant classified phyla (> 1 %) by relative

sequence abundance in cows (Chapters 2 and 3) were the Bacteroidetes (~ 43 %), Firmicutes (~ 40 %), Proteobacteria (~ 4.2 %), and Tenericutes (~ 3.3 %). The fifteen West African Dwarf (WAD) sheep ewes on the control diet of *P. maximum* grass (Chapter 4) were dominated by Firmicutes (58 %), followed by Bacteroidetes (19 %), Tenericutes (6.6 %), Actinobacteria (3.2 %), and Proteobacteria (2.4 %). Actinobacteria have previously been found in dairy cattle rumens at low abundance in large-scale sequencing studies [4,5,7] ,were 0.79 % of sequences found in a cross-study comparison of ruminal sequences [2], and were a minor component in the cows used for this dissertation at 0.07 % in both bovine rumen studies. To date there are only two large-scale sequencing studies published outside of our work on the sheep rumen, of which one used a single wether in which no Actinobacteria were detected [13], and the other found Actinobacteria at extremely minor (0.004 %) relative sequence abundance in five ewes' combined ruminal solids [14]. Previously work with the sheep rumen detected Actinobacteria at 0.75 % of clones in a clone library [15] and as 1.4 % of the cultures from a single wether [16]. Based on these combined results it appears that the Actinobacteria are highly variable members of the ruminal community, but may be more prominent in sheep rather than in cattle.

Although total community composition was easily and significantly separable by ruminal phase of origin (liquids *versus* solids, for an example see Chapter 2, Fig. 1) in our work there was little consistency in phase-specific membership. It has previously and repeatedly been shown that the bacteria present in the liquid and solid fractions of the rumen are overlapping but not identical communities [4,10,17], but the lists of phase-associated bacteria vary widely by study and are probably diet, species, and environment-specific. I found, using multiple methods of identifying potentially important bacteria (Chapters 2 and 3), that the separation between ruminal solid and liquid bacterial communities was primarily due to differences in the liquid-

associated Prevotellaceae (particularly the genus *Prevotella*) and the solid-associated Lachnospiraceae (particularly the genera *Lachnospira* and *Coprococcus*). Fouts *et al.* similarly reported that the liquid phase of cows was associated with *Prevotella*, while *Butyrivibrio* and *Blautia* (in the Lachcnospiraceae) were associated with the solid phase, but additionally reported four other OTUs not seen in my work [17]. The OTUs responsible for differences between the phase-associated communities may be inconsistent between studies due to the transitory presence of diet-derived bacteria passing through the rumen in the liquid phase.

At the genus level, *Butyrivibrio*, *Prevotella*, and *Ruminococcus* are considered "predominant" ruminal bacteria [18], and were three of the four genera, along with *Coprococcus*, present in every animal and study from this dissertation (Table 1). In all cases the genus with the greatest relative sequence abundance, by roughly 10-fold, was *Prevotella*. These four shared genera, among others, have previously been reported as "core" ruminal bacteria in lactating dairy cows [5] and match four of the five most abundant OTUs found by Szpakowski *et al.* among 12 cows of multiple breeds [17]. Species of *Butyrivibrio* are starch, hemicellulose, pectin, and cellulose degraders, producing butyrate, formate, and acetate [18]. *Coprococcus* is capable of using the plant flavonoid component phloroglucinol as a sole carbon source [19], is commonly reported from cattle rumens [2,17,20,21,22], and generates butyrate, acetate, formate, lactate, and propionate from glucose [23]. Other genera traditionally associated with the rumen, such as *Fibrobacter*, *Lachnospira*, *Ruminobacter*, and *Selenomonas*, were not consistently detected in the work reported here or elsewhere; this variation may be influenced by the loss of DNA during physical disruption steps of the extraction process and the difficulty in amplifying even purified DNA from some bacteria [13].

Table 1. Relative sequence abundance of ruminal genera shared by all animals in at least one study when solid and liquid phase sequence sets are combined.

	Relative sequence abundance (%)		
Genus	**Cow 1** [a]	**Cow 2**	**WAD sheep**
Prevotella	29.51	40.08	8.94
Butyrivibrio	3.58	2.38	6.28
Ruminococcus	2.82	2.35	1.38
Coprococcus	2.52	2.29	0.89
Lachnospira	1.44	0.80	-[c]
Pseudobutyrivibrio	1.08	0.81	ND
Treponema	0.77	0.79	-
Paludibacter	0.69	ND	-
Selenomonas	0.65	-	0.45
Moryella	0.64	-	-
Fibrobacter	0.63	0.48	-
Ruminobacter	0.52	-	-
Succinivibrio	0.47	0.38	-
Eubacterium	0.46	-	1.13
Anaeroplasma	0.34	-	-
RFN20 (Erysipelotrichaceae)	0.28	-	-
p-75-a5 (Erysipelotrichaceae)	0.22	-	-
Oscillospira	0.21	-	-
Shuttleworthia	0.18	1.10	-
L7A E11 (Erysipelotrichaceae)	0.16	-	-
Sutterella	0.09	-	ND
Pyramidobacter	0.08	-	-
SHD-231 (Anaerolinaceae)	0.06	-	1.32
Dehalobacterium	0.05	-	-
Clostridium	0.04	1.74	0.41
Bulleidia	0.03	ND	-
Lachnobacterium	0.02	ND	-
Succiniclasticum	ND[b]	2.28	ND
CF231 (Paraprevotellaceae)	ND	1.25	ND
YRC22 (Paraprevotellaceae)	ND	1.04	ND
Acetivibrio	-	ND	0.46

[a] Cow 1 = Chapter 2, eight cows; Cow 2 = Chapter 3, 14 cows; WAD sheep = Chapter 4, seven West African Dwarf sheep

[b] ND = the genus did not have representation in any animal within that study

[c] - = the genus did not have representation in all animals within that study

In summary, previous studies and the work described in this dissertation indicate that ruminal bacterial communities are highly variable, sharing only a few core members across individuals and species. Despite the diversity in bacterial populations the biochemical functions of the rumen are similar. Thus, it is probable that functional groups of bacteria, rather than specific species or strains, drive the bulk of the ruminal community composition as has been proposed to explain the wide diversity of the human gut bacterial community [24].

Milk production efficiency correlates with ruminal bacterial community composition

There is a growing understanding that ruminant milk production [12,25], weight gain [26,27,28], and health [8,29,30] are all linked to the membership of the ruminal microbial community. In this dissertation I focused milk production as an aspect of feed conversion efficiency, measuring the quantity and quality of milk against the feed consumption and energy requirements of lactating dairy cows (Chapters 2 and 3). Milk production is dependent upon ruminal function in that the synthesis of many milk components uses fermentation products released during microbial digestion of feed in the rumen [31,32,33]. It is thus possible that a cow creating large volumes of high-fat, high-protein milk while consuming minimal quantities of feed possesses a microbial community in her rumen distinct from that present in low efficiency animals, and that her relatively higher production efficiency is in large part due to that community difference. I tested for a correlation of the bacterial community with host efficiency in two studies (Chapters 2 and 3), and in both found that there was a correlation.

In the first study (Chapter 2, PJW09) eight cows were assigned to relative efficiency pairs in order to account for differences in parity (the number of times a cow has calved) and lactation stage, and were tracked for a relatively short period of time (nine days). In the second study

(Chapter 3, PJW10) I tracked 14 cows over two lactation cycles and compared them at identical parity and lactation stage, making it possible to treat gross feed efficiency (GFE) as the continuous variable. In both studies there were strong and significant correlations between the total ruminal bacterial community composition and host efficiency. This pattern was visible for the entire ruminal community (Chapter 3, Fig. 6A-C) and for each of the individual ruminal phases (Chapter 3, Fig. 7), indicating that the degree of correlation observed was not due to either the liquid or solid phase alone. This importance of the total ruminal bacterial community, and not just of one phase, is in agreement with the results of our preliminary cow efficiency study (Chapter 2), where multiple OTUs in both the liquid and solid phases were found to strongly correlate with relative efficiency (Table 3.

Both studies found multiple *Prevotella* OTUs correlating with either increased or decreased efficiency, suggesting that different species or strains of this genus have opposing impacts on the total host efficiency. As *Prevotella* species in the rumen are known to ferment starch, pectin, xylan, and sugars to succinate, acetate, formate, and propionate [18] it is possible that differential production of these fermentation end-products significantly shifts the total volatile fatty acids (VFAs) available for host uptake and metabolism. Similarly, there were OTUs classified as Lachnospiracea in both the higher and lower efficiency correlation sets, and many genera within the Lachnospiraceae associated with increased (*Shuttleworthia* and *Butyrivibrio*) and decreased (*Butyrivibrio* and *Lachnospira*) efficiency. There is no clear pattern or link among the OTU or family associations, suggesting that the importance of any species may be due to specific metabolic capabilities or interactions with other community members.

These results give rise to two competing hypotheses that remain to be tested. The first is that it is the total functional bacterial community, and not the rise or decline of a specific OTU,

that has the greatest degree of influence on host efficiency. If true, then differences in the number and diversity of genes, such as those responsible for cellulase degrading enzymes, will correlate highly with efficiency and be detectable by comparing ruminal metagenomes. Such differences in ruminal microbial gene expression would be measurable using techniques such as RNA-seq. To date the focus on ruminal metagenomes is for biofuel-related projects [34,35,36], although the feasibility of metagenomic profiling for studies of host parameters has been proven [37] and is beginning to be used for cattle studies incorporating host factors [38,39] by the group that proposed the technique. Because these ruminal metagenomes are not sorted into individual species, relying instead on nonspecifically amplified whole-community genomic profiles, it is difficult to compare our 16S rRNA gene-specific results with metagenomes. An alternate hypothesis is that individual and specific bacterial species have a measurable impact on the host cow's milk production efficiency. To address this hypothesis, work should be done using quantitative PCR to enumerate and more clearly identify the OTUs shown by correlation analyses in this dissertation to match with high and low efficiency animals. Because both the metagenomic [39] and gene-specific methods (this dissertation and [12]) have shown correlations between the ruminal community and host production parameters it is not yet possible to say whether function or taxonomy has the greater contribution to host efficiency, or a correct efficiency model must incorporate both ideas.

The conclusions made in this thesis would be improved by adding diversity and population information for ruminal fungi and methanogens. Anaerobic fungi in the Neocallimastigomycota, as primary cellulose degraders, have been shown in a dosing experiment with water buffalo to have a direct and positive impact on milk production and milk fat content [25]. Methanogens in the rumen act as hydrogen sinks through the reduction of CO_2 to methane,

but are not considered essential for rumen function, and much research has been published on efforts to remove or diminish methanogen populations (for a review, see [40]). Although highly variable depending upon diet and methanogen population size, host-available energy lost to methane production can range from 2 – 7 % [41], and it is possible that specific populations of methanogens, like bacteria, correlate with feed conversion and milk production efficiency.

Diet has a strong but idiosyncratic impact on ruminal bacteria

It would be difficult to list all of the studies published demonstrating the significant impact of diet on the ruminal bacterial community. Feed additives for sheep and cattle range from TNT [15] and antibiotics [42] to fish oil [43] and essential oils [44], but the major dietary components are starches from grain and fiber from feed such as alfalfa or grass. In brief, increasing proportions of feedstock result in larger populations of fiber-degrading bacteria in the Lachnospiraceae, Ruminococcaceae, and Fibrobacteraceae [9,45,46,47], while increasing proportions of grain encourages the growth of starch-degrading bacteria in the Flavobacteriaceae, Prevotellaceae, and Veillonellaceae [9,30,47,48]. Even within this broad generalization, however, there is a great deal of variation. For example, Fernando *et al.* (2010) found by qPCR that high levels of grain in place of hay in the diet of cattle increased the ruminal populations of *Streptococcus bovis* (Streptococcaceae) and *Prevotella bryantii* [46] (Prevotellaceae), while Tajima *et al.*, also using qPCR and diet switching in cattle, reported that *S. bovis* and *P. bryantii* populations rose and then fell to levels similar to pre-grain supplementation during diet adaptation [45]. In both studies *Ruminococcus flavefaciens* (Ruminococcaceae) populations dropped with increasing grain supplementation [45,46], but work by Petri *et al.* showed, also by qPCR, that *Ruminococcus* spp. populations were equal between cattle on a high concentrate diet

with or without forage [47]. These, and other comparisons between diet-shift studies in cattle and sheep, highlight the variability in the presence and/or detection of bacterial populations.

The ruminant diet does not require grain, and especially in developing countries there is an emphasis on animal diets relying on non-agricultural, native plants providing adequate nutrition and not negatively impacting rumen function [49,50,51,52]. In our diet-change study (Chapter 4) with WAD sheep in Nigeria we supplemented a basal, high-fiber diet of the grass *Panicum maximum* with leaves and thin branches from the local trees *Albizia saman*, *Bridelia micrantha, Ficus sur,* or *Gmelina arborea*. We concluded from the study that two of the trees, *A. saman* and *G. arborea*, caused major shifts in the ruminal bacterial community, but we were unable to account for these shifts by comparisons of the known feed composition profiles (crude protein, hemicellulose, tannin, phytate, *etc*.). It is possible that dietary impact was due to a non-nutritive small molecule that was directly inhibitory to the growth of specific bacteria. This study also highlighted the difficulty in pooling animals, in that every ewe in the study, even on the control diet, had a distinct and highly dissimilar community from each of the other ewes (Chapter 4, Fig. 1A & B). Because of this high level of community difference within each feeding trial we were able to tell that the communities changed, but not draw strong conclusions regarding the impact of diet on specific OTUs.

Impact of time on the ruminal bacterial community

Many studies of feed efficiency and/or ruminal community implicitly assume that both are stable when all other variables are held constant in comparisons between animals of mixed (or unreported) age or days in milk (DIM) [4,5,9,11,37,53,54,55,56,57,58,59,60,61]. The inability to know the age of an animal is understandable in studies relying on wild-caught

animals [11,61], but the potential impact of age on their results must be acknowledged. The frequent lack of age or DIM inclusion in studies is particularly surprising given the long-known physiological changes that occur in dairy cows over the course of each lactation cycle [62]. The work presented in Chapter 3 of this dissertation indicates significant flaws in the assumption of similarity among animals of different age, and especially of different DIM, when comparing ruminal microbial communities.

When performing our preliminary cow efficiency study (Chapter 2) we assumed that lactation period could be a significant variable and included a term for DIM in the relative efficiency calculations (Table 1). The eight cows were in their middle-to-late lactation periods (mean of 218 ± 47 DIM), and when compared to the later study were similar ($P = 0.413$, Student's T-test) in terms of their inverse Simpson's diversity (110 ± 17 for the eight cows to 104 ± 20 for the 14 cows). The lower number of observed unique OTUs in the pooled Chapter 3 sequence sets (610 ± 107 *vs.* 996 ± 53 unique OTUs) was most likely due to the significantly ($P < 0.00001$, Student's T-test) lower sequence coverage of Chapter 3 (mean of 92.28 %) as compared to Chapter 2 (mean of 97.09 %).

First, the importance of only comparing animals of similar lactation stage was demonstrated by the gross shifts in community composition found as lactation progressed from Early to Late periods (Chapter 3, Fig. 5). Similarly, comparisons of efficiency were period-dependent ($P < 0.0001$, repeated measures ANOVA with matched observations). It is possible that the importance of only comparing within similar sampling periods is due to the physiological shifts in immunological state and energy utilization [62,63] during a cow's lactation cycle. So long as lactation stage was held constant the total ruminal community compositions were indistinguishable by lactation cycles (Chapter 3, Fig. 2 and Fig. 4),

suggesting that the impact of age on the total bacterial community composition is less important than lactation stage. An additional variable, untested our work, is the impact of temperature on the ruminal community. A previous study by Uyeno *et al.* (2010) demonstrated that heat stress has the potential to measurably change the relative proportion of specific bacterial taxa, such as decrease in *Fibrobacter* and increase of *Streptococcus*, with increasing heat [64]. In that work heifers were maintained in temperature-controlled environments, but it is probable that the vast differences between the summer and winter barn temperatures similarly influenced our measured ruminal communities. The barns used in our two cow studies were cooled in summer by fans and breeze-ways to avoid heat accumulation, and heated in winter by cow body heat to near-freezing to avoid the freezing of trough water, but were otherwise unregulated in terms of temperature. As the barn temperatures were not recorded in our own work we cannot further analyze the impact of heat or cold stress in our own data sets.

Second, although the total bacterial community composition did not significantly change between the first and second lactation cycles, the inverse Simpson's diversity continued to increase in the liquid fraction as the cows aged (Chapter 3). Similarly, there was a trend towards increasing diversity in the pooled (ruminal solids plus liquids) diversity, from 36.44 – 141.34 in the first lactation to 49.55 – 145.78 in the second lactation; for the short-term cow study (Chapter 2), where all eight of the cows were in their second or third lactation, the range was from 82.27 – 134.41. By these values it appears that bacterial diversity reaches its maximum and plateaus between the second and third lactation cycle at 3-4 years of age, assuming the cow was successfully impregnated each year. It is possible that extended physical maturation plays a large role in determining the ruminal diversity. This may in part explain the surprisingly low diversity range of 3.51 – 45.45 detected in the WAD sheep (Chapter 4), which were all under one

year of age. The combined results of these three studies suggest that ruminal diversity continues to increase after sexual maturity, but steadies as the animal reaches full body maturity.

References

1. van Dijk EL, Jaszczyszyn Y, Thermes C (2014) Library preparation methods for next-generation sequencing: Tone down the bias. Experimental Cell Research 322: 12-20.
2. Kim M, Morrison M, Yu Z (2011) Status of the phylogenetic diversity census of ruminal microbiomes. FEMS Microbiology Ecology 76: 49-63.
3. Brulc JM, Antonopoulos DA, Berg Miller ME, Wilson MK, Yannarell AC, et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proceedings of the National Academy of Sciences 106: 1948-1953.
4. de Menezes AB, Lewis E, O'Donovan M, O'Neill BF, Clipson N, et al. (2011) Microbiome analysis of dairy cows fed pasture or total mixed ration diets. FEMS Microbiology Ecology 78: 256-265.
5. Jami E, Mizrahi I (2012) Composition and similarity of bovine rumen microbiota across individual animals. PLOS ONE 7: e33306.
6. Sandri M, Manfrin C, Pallavicini A, Stefanon B (2014) Microbial biodiversity of the liquid fraction of rumen content from lactating cows. Animal 8: 572-579.
7. Ishaq S, Wright A-D (2014) High-throughput DNA sequencing of the ruminal bacteria from moose (*Alces alces*) in Vermont, Alaska, and Norway. Microbial Ecology: 1-11.
8. Mao SY, Zhang RY, Wang DS, Zhu WY (2013) Impact of subacute ruminal acidosis (SARA) adaptation on rumen microbiota in dairy cattle using pyrosequencing. Anaerobe 24: 12-19.
9. Thoetkiattikul H, Mhuantong W, Laothanachareon T, Tangphatsornruang S, Pattarajinda V, et al. (2013) Comparative Analysis of Microbial Profiles in Cow Rumen Fed with Different Dietary Fiber by Tagged 16S rRNA Gene Pyrosequencing. Current Microbiology 67: 130-137.
10. Cunha IS, Barreto CC, Costa OYA, Bomfim MA, Castro AP, et al. (2011) Bacteria and Archaea community structure in the rumen microbiome of goats (*Capra hircus*) from the semiarid region of Brazil. Anaerobe 17: 118-124.
11. Gruninger RJ, Sensen CW, McAllister TA, Forster RJ (2014) Diversity of Rumen Bacteria in Canadian Cervids. PLOS ONE 9: e89682.
12. Jami E, White BA, Mizrahi I (2014) Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. PLOS ONE 9: e85423.
13. Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, et al. (2013) Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. PLOS ONE 8: e74787.
14. Castro-Carrera T, Toral PG, Frutos P, McEwan NR, Hervás G, et al. (2014) Rumen bacterial community evaluated by 454 pyrosequencing and terminal restriction fragment length polymorphism analyses in dairy sheep fed marine algae. Journal of Dairy Science 97: 1661-1669.
15. Perumbakkam S, Mitchell EA, Morrie Craig A (2011) Changes to the rumen bacterial population of sheep with the addition of 2,4,6-trinitrotoluene to their diet. Antonie van Leeuwenhoek Journal of Microbiology 99: 231-240.
16. Kenters N, Henderson G, Jeyanathan J, Kittelmann S, Janssen P (2011) Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. Journal of Microbiological Methods 84: 52-60.

17. Fouts D, Szpakowski S, Purushe J, Torralba M, Waterman R, et al. (2012) Next generation sequencing to define prokaryotic and fungal diversity in the bovine rumen. PLOS ONE 7: e48289.
18. Russell JB, Rychlik J (2001) Factors that alter rumen microbial ecology. Science 292: 1119-1122.
19. Patel TR, Jure KG, Jones GA (1981) Catabolism of Phloroglucinol by the Rumen Anaerobe Coprococcus. Applied and Environmental Microbiology 42: 1010-1017.
20. Jami E, Israel A, Kotser A, Mizrahi I (2013) Exploring the bovine rumen bacterial community from birth to adulthood. ISME Journal 7: 1069-1079.
21. Li RW, Connor EE, Li C, Baldwin VI RL, Sparks ME (2012) Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. Environmental Microbiology 14: 129-139.
22. Hernandez-Sanabria E, Guan LL, Goonewardene LA, Li M, Mujibi DF, et al. (2010) Correlation of Particular Bacterial PCR-Denaturing Gradient Gel Electrophoresis Patterns with Bovine Ruminal Fermentation Parameters and Feed Efficiency Traits. Applied and Environmental Microbiology 76: 6338-6350.
23. Holdeman LV, Moore WEC (1974) New genus, *Coprococcus*, twelve new species, and emended descriptions of four previously described species of bacteria from human feces. International Journal of Systematic Bacteriology 24: 260-277.
24. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, et al. (2011) Enterotypes of the human gut microbiome. Nature 473: 174-180.
25. Saxena S, Sehgal J, Puniya A, Singh K (2010) Effect of administration of rumen fungi on production performance of lactating buffaloes. Beneficial Microbes 1: 183-188.
26. Guan LL, Nkrumah JD, Basarab JA, Moore SS (2008) Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. FEMS Microbiology Letters 288: 85-91.
27. Hernandez-Sanabria E, Goonewardene LA, Wang Z, Durunna ON, Moore SS, et al. (2012) Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. Applied and Environmental Microbiology 78: 1203-1214.
28. Carberry CA, Kenny DA, Han S, McCabe MS, Waters SM (2012) Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. Applied and Environmental Microbiology 78: 4949-4958.
29. Hook SE, Steele MA, Northwood KS, Dijkstra J, France J, et al. (2011) Impact of subacute ruminal acidosis (SARA) adaptation and recovery on the density and diversity of bacteria in the rumen of dairy cows. FEMS Microbiology Ecology 78: 275-284.
30. Khafipour E, Li S, Plaizier JC, Krause DO (2009) Rumen microbiome composition determined using two nutritional models of subacute ruminal acidosis. Applied and Environmental Microbiology 75: 7115-7124.
31. Fievez V, Colman E, Castro-Montoya JM, Stefanov I, Vlaeminck B (2012) Milk odd- and branched-chain fatty acids as biomarkers of rumen function—An update. Animal Feed Science and Technology 172: 51-65.
32. Vlaeminck B, Fievez V, Cabrita ARJ, Fonseca AJM, Dewhurst RJ (2006) Factors affecting odd- and branched-chain fatty acids in milk: A review. Animal Feed Science and Technology 131: 389-417.
33. Bauman D, Griinari J (2003) Nutritional regulation of milk fat synthesis. Annual Review of Nutrition 23: 203-227.

34. Gong X, Gruniniger RJ, Forster RJ, Teather RM, McAllister TA (2013) Biochemical analysis of a highly specific, pH stable xylanase gene identified from a bovine rumen-derived metagenomic library. Applied Microbiology and Biotechnology 97: 2423-2431.
35. Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, et al. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331: 463-467.
36. Del Pozo MV, Fernández-Arrojo L, Gil-Martínez J, Montesinos A, Chernikova TN, et al. (2012) Microbial β-glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. Biotechnology for Biofuels 5: doi:10.1186/1754-6834-1185-1173.
37. Ross E, Moate P, Bath C, Davidson S, Sawbridge T, et al. (2012) High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. BMC Genetics 13: doi:10.1186/1471-2156-1113-1153.
38. Ross EM, Moate PJ, Marett L, Cocks BG, Hayes BJ (2013) Investigating the effect of two methane-mitigating diets on the rumen microbiome using massively parallel sequencing. Journal of Dairy Science 96: 6030-6046.
39. Ross EM, Moate PJ, Marett LC, Cocks BG, Hayes BJ (2013) Metagenomic Predictions: From Microbiome to Complex Health and Environmental Phenotypes in Humans and Cattle. PLoS ONE 8: e73056.
40. McAllister T, Newbold C (2008) Redirecting rumen fermentation to reduce methanogenesis. Australian Journal of Experimental Agriculture 48: 7-13.
41. Johnson K, Johnson D (1995) Methane emissions from cattle. Journal of Animal Science 73: 2483-2492.
42. Weimer PJ, Stevenson DM, Mertens DR, Thomas EE (2008) Effect of monensin feeding and withdrawal on populations of individual bacterial species in the rumen of lactating dairy cows fed high-starch rations. Applied and Environmental Microbiology 80: 135-145.
43. Liu S, Bu D, Wang J, Liu L, Liang S, et al. (2011) Effect of incremental levels of fish oil supplementation on specific bacterial populations in bovine ruminal fluid. Journal of Animal Physiology and Animal Nutrition DOI: 10.1111/j.1439-0396.2010.01113.x.
44. Benchaar C, Petit H, Berthiaume R, Ouellet D, Chiquette J, et al. (2007) Effects of essential oils on digestion, ruminal fermentation, rumen microbrial populations, milk production, and milk composition in diary cows fed alfalfa silage or corn silage. Journal of Dairy Science 90: 886-897.
45. Tajima K, Aminov R, Nagamine T, Matsui H, Nakamura M, et al. (2001) Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR. Applied and Environmental Microbiology 67: 2766-2774.
46. Fernando S, Purvis II H, Najar F, Sukiharnikov L, Krehbiel C, et al. (2010) Rumen microbial population dynamics during adaptation to a high-grain diet. Applied and Environmental Microbiology 76: 7482-7490.
47. Petri R, Forster R, Yang W, McKinnon J, McAllister T (2012) Characterization of rumen bacterial diversity and fermentation parameters in concentrate fed cattle with and without forage. Journal of Applied Microbiology 112: 1152-1162.
48. Counotte GHM, Prins RA, Janssen RHAM, deBie MJA (1981) Role of *Megasphaera elsdenii* in the fermentation of dl-[2-^{13}C]lactate in the rumen of dairy cattle. Applied and Environmental Microbiology 42: 649-655.

49. Ouédraogo-Koné S, Kaboré-Zoungrana C, Ledin I (2008) Intake and digestibility in sheep and chemical composition during different seasons of some West African browse species. Tropical Animal Health and Production 40: 155-164.
50. Ammar H, López S, González JS, Ranilla MJ (2004) Chemical composition and in vitro digestibility of some Spanish browse plant species. Journal of the Science of Food and Agriculture 84: 197-204.
51. Aletor V, Omodara O (1994) Studies on some leguminous browse plants with particular reference to their proximate, mineral and some endogenous anti-nutritional constituents. Animal Feed Science and Technology 46: 343-348.
52. Le Houérou HN (1980) Chemical composition and nutritive value of browse in tropical West Africa. In: Le Houérou HN, editor. International Symposium on Browse in Africa: Internation Livestock Centre for Africa. pp. 261-289.
53. Lee J, Jung J, Oh Y, Lee S, Madsen E, et al. (2012) Comparative survey of rumen microbial communities and metabolites across on caprine and three bovine groups, using bar-coded pyrosequencing and ^{1}H nuclear magnetic resonance spectroscopy. Applied and Environmental Microbiology 78: 5983-5993.
54. Kang SH, Evans P, Morrison M, McSweeney C (2013) Identification of metabolically active proteobacterial and archaeal communities in the rumen by DNA- and RNA-derived 16S rRNA gene. Journal of Applied Microbiology 115: 644-653.
55. Lettat A, Benchaar C (2013) Diet-induced alterations in total and metabolically active microbes within the rumen of dairy cows. PLOS ONE 8: e60978.
56. Kittelmann S, Seedorf H, Walters W, Clemente JC, Knight R, et al. (2013) Simultaneous amplicon sequencing to explore co-occurence patterns of bacterial, archaeal, and eukaryotic microorganisms in rumen microbial communities. PLOS ONE 8: e47879.
57. Mullins CR, Mamedova LK, Carpenter AJ, Ying Y, Allen MS, et al. (2013) Analysis of rumen microbial populations in lactating dairy cattle fed diets varying in carbohydrate profiles and Saccharomyces cerevisiae fermentation product. Journal of Dairy Science 96: 5872-5881.
58. Saro C, Ranilla MJ, Carro MD (2012) Postprandial changes of fiber-degrading microbes in the rumen of sheep fed diets varying in type of forage as monitored by real-time PCR and automated ribosomal intergenic spacer analysis. Journal of Animal Science 90: 4487-4494.
59. Kim M, Yu Z (2012) Quantitative comparisons of select cultured and uncultured microbial populations in the rumen of cattle fed different diets. Journal of Animal Science and Biotechnology 3: doi:10.1186/2049-1891-1183-1128.
60. Sundset M, Edwards J, Cheng Y, Senosiain R, Fraile M, et al. (2009) Molecular diversity of the rumen microbiome of Norwegian reindeer on natural summer pasture. Microbial Ecology 57: 335-348.
61. Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, et al. (2012) Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. PLOS ONE.
62. National Research Council (2001) Nutrient requirements of dairy cattle: National Academy Press, Washington DC.
63. Goff JP, Horst RL (1997) Physiological Changes at Parturition and Their Relationship to Metabolic Disorders1,2. Journal of Dairy Science 80: 1260-1268.

64. Uyeno Y, Sekiguchi Y, Tajima K, Takenaka A, Kurihara M, et al. (2010) An rRNA-based analysis for evaluating the effect of heat stress on the rumen microbial composition of Holstein heifers. Anaerobe 16: 27-33.

APPENDICES

APPENDIX ONE

A phylogenetic analysis of the phylum *Fibrobacteres*

Originally published as:

Jewell, Kelsea A., Jarrod J. Scott, Sandye M. Adams, and Garret Suen. (2013) A phylogenetic analysis of the phylum *Fibrobacteres*. *Systematic and Applied Microbiology*. 36(6): 376-382.

Modifications have been made to Table S1 to increase legibility in the current format and to the references for consistency in this thesis.

KJ, JS, and SA created the sequence database. KJ performed the sequence analyses and wrote the manuscript. GS designed the experiments and assisted with writing the manuscript.

Abstract

Members of the phylum *Fibrobacteres* are highly efficient cellulolytic bacteria, best known for their role in rumen function and as potential sources of novel enzymes for bioenergy applications. Despite being key members of ruminal and other digestive microbial communities, our knowledge of this phylum remains incomplete, as much of our understanding is focused on two recognized species, *Fibrobacter succinogenes* and *F. intestinalis*. As a result, we lack insights regarding the environmental niche, host range, and phylogenetic organization of this phylum. Here, we analyzed over 1,000 16S rRNA *Fibrobacteres* sequences available from public databases to establish a phylogenetic framework for this phylum. We identify both species- and genus-level clades that are suggestive of previously unknown taxonomic relationships between *Fibrobacteres* in addition to their putative lifestyles as host-associated or free-living. Our results shed light on this poorly understood phylum and will be useful for elucidating the function, distribution, and diversity of these bacteria in their niches.

Introduction

The ability of herbivores to convert plant biomass into usable nutrients is predicated on symbiotic associations with diverse microbial communities [1]. A key example is ruminants, which use a consortium of microbes to degrade and ferment recalcitrant forms of cellulosic biomass into short-chain, host-available volatile fatty acids (for a review, see [2]). Among these important microbes is the bacterium *Fibrobacter succinogenes* [3], a prolific cellulose degrader [4,5] that produces succinic acid as its major fermentation product and lesser amounts of acetic and formic acids. The recently completed genome sequence for the type strain, *F. succinogenes* S85, highlighted the metabolic and cellulolytic specialization of this bacterium in the rumen [6]. *F. succinogenes* belongs to the phylum *Fibrobacteres*, and work within this phylum has focused primarily on this species and *F. intestinalis*, which is typically found associated with non-ruminant mammalian guts [7]. As a result, our knowledge of the *Fibrobacteres*' host range, environmental niche, and species diversity is based almost entirely on *F. succinogenes* and (to a much lesser extent) *F. intestinalis*. A better understanding of this phylum, and its member species, will be useful for guiding research in areas such as agriculture and biofuel production.

The genus *Fibrobacter* was previously classified within the *Bacteroidetes*, but was elevated to its own phylum based on 16S rRNA sequence analysis and physiological differences from other members of the *Bacteroidetes* [7]. The *Fibrobacteres* are currently defined as anaerobic, gram-negative, non-spore forming, cellulolytic, non-motile rods [7]. However, only two *Fibrobacter* species have been cultured and formally described (*F. succinogenes* and *F. intestinalis*) [7], despite 16S rRNA sequence data that strongly suggests that cryptic species exist [8,9,10,11]. At present, our understanding of *Fibrobacteres* physiology is based entirely upon *F. succinogenes* and *F. intestinalis* isolates associated with animals [10,11,12,13,14]. These studies

do not reflect the wide diversity of 16S rRNA sequences identified as *Fibrobacteres* that have been reported from surveys of environments as disparate as landfills [12], freshwater lakes [13], the ocean [14], limestone cave sulfidic waters [15], termite hindguts [16,17], and a wide variety of animal feces [18,19]. Given the paucity of information available for the *Fibrobacteres*, and the highly desirable cellulolytic properties of its currently described members, it is important to establish a knowledgebase that documents the diversity of species, distribution, and host associations that exist within this phylum.

Here, we capitalize on extensive public databases of existing 16S rRNA sequence data to create and analyze a *Fibrobacteres*-specific phylogeny. This dataset includes many sequences of environmental origin, in addition to host-associated sequences, and we use this data to estimate the diversity present in non-ruminant and non-fecal *Fibrobacteres*. Our analysis reveals a diverse phylogeny that we use to estimate both the species structure and environmental distribution of this poorly understood phylum.

Materials and methods

Sequence data collection, screening, and phylogenetic analysis. All 16S rRNA sequences available through the National Center for Biotechnological Information's (NCBI) nucleotide database marked with the search terms "*Fibrobacteres*," "Fibrobacter," or "Fibrobact*" (where "*" indicates a wild-card search term, accessed: 09/17/2012) were used to construct an initial sequence library. Roche 454-based pyrosequence libraries were not included, as their short average read length complicates the ability to generate usable alignments for downstream analyses [20]. Literature searches were conducted to identify other 16S rRNA sequences likely belonging to the phyla "*Fibrobacteres*" or "*Fibrobacteres*/*Acidobacteria*" but not marked as

such in NCBI. The *Fibrobacteres* sequences present in the GreenGenes [21], Ribosomal Database Project [22] and Silva [23] sequence repositories were also included. The total sequence set was annotated to include sequence source and location from both GenBank deposit information and the original publication, if such existed. The complete dataset is presented in Table S1.

The following phylogenetic analysis was performed on the compiled sequence dataset. All sequences were imported into ARB [24] and a full alignment was created against the current Silva 16S/18S rRNA non-redundant sequence database (SSU Ref NR; release 102; 262,092 total sequences); sequences with closer affinity to known *Fibrobacteres* than to any other phyla were considered as belonging to the phylum *Fibrobacteres*. The *Fibrobacteres*-associated sequences were processed in mothur (v.1.26.0, commands used in the following description denoted in italics) [25]. Sequences > 900 bp with 5 or fewer ambiguous nucleotides and nine or fewer homopolymers were retained (*screen.seqs*), duplicate sequences were removed (*unique.seqs*), and a preliminary alignment was created (*align.seqs*, Needleman-Wunsch pairwise alignment method, gap extension penalty = -1, gap opening penalty = -1, and match = +1, mismatch penalty = -1, k size = 7). Chimera detection (*chimera.uchime*) used the Silva 16S/18S rRNA non-redundant sequence database (SSU Ref NR, accessed 09/2012). Aligned sequences and were filtered (*filter.seqs*, trump = ., vertical = T, soft = 50) and used to create a distance matrix (*dist.seqs*, calc = onegap, cutoff = 0.2). The distance matrix used to calculate the estimated number of operational taxonomic units (OTUs) present at 90 %, 95 %, and 97 % similarity cut-offs (*cluster*, nearest neighbor algorithm). Representative sequences were chosen for each OTU (*get.oturep*) for use in constructing a tree in MrBayes (v3.1) [26,27] (ngen = 10,000,000, chain = 4), with the resulting tree visualized using FigTree (v1.3.1) [28]. In addition, a Neighbor-Joining

tree was generated from all sequences [29] using the Maximum Composite Likelihood method [30] in MEGA5 [31] and visualized using the Interactive Tree of Life project [32]. The complete 16S rRNA sequence for *Bacteroides fragilis* NCTC9343 was included in our phylogenetic analyses as an outgroup (GenBank genome accession number: NC_003228.3).

A second Neighbor-Joining tree was constructed for sequences of 450 bp or greater in length using these same methods. Some modifications to the sequence manipulations were required due to this shortened minimal sequence length, and were as follows: no ambiguous nucleotides were allowed, and the alignment filter did not include *trump* = .. After alignment and filtering the dataset was used to create a p-distance pair-wise distance matrix in MEGA5, and those sequences for which it was not possible to estimate evolutionary distances were removed.

Results

Distribution and definition of the *Fibrobacteres*. From an initial database of 1,166 putative *Fibrobacteres* sequences we generated a database of 863 confirmed *Fibrobacteres* sequences of > 900 bp (henceforth, "long sequence database") and 1,095 sequences > 450 bp (henceforth, "short sequence database") after all filtering steps were performed (Table S1 and Fig. S1). The 900 bp length cut-off was chosen to retain a maximum number (at least 70%) of the original sequences while minimizing the effects of reduced alignment quality (number of columns removed during alignment filtering). The shorter *Fibrobacteres* database was constructed for comparative purposes. The results presented here, except when stated otherwise, refer to analyses performed using the long sequence database.

Table S1. Summary of all sequences initially included in the *Fibrobacteres* phylogenetic analyses.

GenBank ID range	Reason for removal	Generalized host	Source or strain ID	Citation	97% OTU	95% OTU
AB275483.1-AB275485.1	Retained	Sheep R [a]	Japan Sapporo Hokkaido	[33]	5	1
AB275486.1-AB275495.1	Retained	Sheep R	Japan Sapporo Hokkaido	[33]	2	1
AB275496.1	Retained	Sheep R	Japan Sapporo Hokkaido	[33]	49	1
AB275497.1-AB275498.1	Retained	Sheep R	Japan Sapporo Hokkaido	[33]	2	1
AB275499.1	Retained	Sheep R	Japan Sapporo Hokkaido	[33]	5	1
AB275500.1-AB275514.1	Retained	Sheep R	Japan Sapporo Hokkaido	[33]	2	1
AF018454.1	Retained	Bovine R	Canada Ottawa	[10]	2	1
AJ496032.1	Retained	Bovine R	*F. succinogenes* S85 ATCC 19169	[4]	2	1
AJ496186.1	Retained	Sheep R	Urbana Illinois	[4]	2	1
AJ496447.1	Retained	Bovine R	*F. succinogenes* H	[4]	2	1
AJ496448.2	Retained	Bovine R	*F. succinogenes* U	[4]	2	1
AJ496566.1	Retained	Sheep R	*F. succinogenes* FE	[4]	2	1
AJ505937.1	Retained	Bovine R	*F. succinogenes* R	[4]	2	1
AJ505938.1	Retained	Bovine R	Canada Alberta	[4]	2	1
EF190826.1	Retained	Landfill	England NW	[12]	2	1
EF190828.1	Retained	Landfill	England NW	[12]	2	1
EF445213.1	Retained	Bovine R	France Tolouse	[34]	2	1
EU381787.1	Retained	Bovine R	Canada Alberta	U [b]	2	1
EU381803.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381811.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381836.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381839.1	Retained	Bovine R	Canada Alberta	U	5	1
EU381840.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381857.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381861.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381922.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381936.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381958.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381968.1	Retained	Bovine R	Canada Alberta	U	2	1
EU381993.1	Retained	Bovine R	Canada Alberta	U	2	1
EU382022.1	Retained	Bovine R	Canada Alberta	U	2	1
EU382049.1	Retained	Bovine R	Canada Alberta	U	2	1
EU463463.1	Retained	horse F	St. Louis Zoo. Park, MO	[19]	64	1
EU463562.1	Retained	horse F	St. Louis Zoo. Park, MO	[19]	65	1
EU468455.1	Retained	Black rhinoceros F	St. Louis Zoo. Park, MO	[19]	5	1
EU470330.1	Retained	Grevys zebra F	St. Louis Zoo. Park, MO	[19]	67	1
EU470375.1	Retained	Grevys zebra F	St. Louis Zoo. Park, MO	[19]	67	1
EU470410.1	Retained	Grevys zebra F	St. Louis Zoo. Park, MO	[19]	67	1
EU473449.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1

EU473476.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU473520.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU473529.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU473538.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU473539.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU473542.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	71	1
EU473545.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	67	1
EU473558.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU473600.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU473604.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU473606.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU475370.1	Retained	rock hyrax F	Zoo. Soc. of San Diego, CA	[19]	2	1
EU475376.1	Retained	rock hyrax F	Zoo. Soc. of San Diego, CA	[19]	2	1
EU606019.1	Retained	Bovine R	India Karnataka	U	2	1
EU774414.1	Retained	Eastern black and white colobus F	St. Louis Zoo. Park, MO	[19]	5	1
EU774452.1	Retained	Eastern black and white colobus F	St. Louis Zoo. Park, MO	[19]	5	1
EU779343.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU779347.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU779383.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU779394.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU779396.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
EU779399.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	70	1
FN429847.1	Retained	*Ventiella sulfuris* tissue	Hydrothermal vent isolate	U	85	1
GQ327172.1	Retained	Bovine R	Canada Alberta	[35]	2	1
GQ358264.1	Retained	Tammar wallaby foregut	Australia Canberra	[36]	5	1
GU269553.1	Retained	Sheep, R	Urbana, IL	[4]	2	1
GU303546.1	Retained	Bovine, R	Canada Alberta	U	2	1
GU303627.1	Retained	Bovine, R	Canada Alberta	U	8	1
GU999988.1	Retained	Goat, R	Malaysia Serdang, Selangor	U	5	1
GU999989.1	Retained	Goat, R	Malaysia Serdang, Selangor	U	5	1
L35548.1	Retained	Pony cecum	Urbana IL	[9]	9	1
M62682.1	Retained	Bovine R	*F. succinogenes* REH9-1	[37]	1	1
M62683.1	Retained	Bovine R	*F. succinogenes* A3C	[37]	2	1
M62684.1	Retained	Bovine R	*F. succinogenes* B1	[37]	2	1
M62685.1	Retained	Bovine R	*F. succinogenes* BL2	[37]	2	1
M62688.1	Retained	Bovine non-R	*F. succinogenes* GC5	[37]	5	1
M62689.1	Retained	Sheep R	*F. succinogenes* HM2	[37]	2	1
M62692.1	Retained	Sheep R	*F. succinogenes* MB4	[37]	7	1
M62693.1	Retained	Sheep R	*F. succinogenes* MCI	[8]	8	1
M62696.1	Retained	Bovine R	*F. succinogenes* S85	[4]	2	1
AJ496284.1	Retained	Rat cecum	Urbana Illinois	[4]	3	2
EU474873.1	Retained	red river hog F	Zoo. Soc. of San Diego, CA	[19]	4	2
EU475285.1	Retained	rock hyrax F	Zoo. Soc. of San Diego, CA	[19]	3	2
EU774496.1	Retained	Eastern black and white colobus F	St. Louis Zoo. Park, MO	[19]	3	2
GQ451204.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2

GQ451231.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451246.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451248.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451260.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451284.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451292.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451306.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451307.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451318.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451324.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
GQ451325.1	Retained	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[38]	3	2
M62686.1	Retained	Pig cecum	*F. intestinalis* C1a	[37]	3	2
M62687.1	Retained	Pig cecum	*F. intestinalis* DR7	[37]	4	2
M62690.1	Retained	Sheep R	*F. intestinalis* JGI	[37]	6	2
M62691.1	Retained	Sheep R	*F. intestinalis* LH1	[37]	6	2
M62695.1	Retained	Rat cecum	*F. intestinalis* NR9	[37]	3	2
L35547.1	Retained	Pony cecum	Urbana Illinois	[9]	10	3
AF165269.1	Retained	Soil	Kohala Forest Reserve HI	[39]	11	4
AB192074.1	Retained	Higher termite G	Thailand Prachinburi	[17]	12	5
AB192075.1	Retained	Higher termite G	Thailand Phitsanulok	[17]	13	5
AB192076.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	12	5
AB192077.2	Retained	Higher termite G	Thailand Pathum Thani	[17]	12	5
AB192089.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	19	5
AB192090.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	19	5
AB192091.1	Retained	Higher termite G	Thailand Nakhon Pathom	[17]	20	5
AB192092.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	20	5
AB192094.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	21	5
AB243277.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	20	5
AB243278.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	12	5
AB255938.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	12	5
AB255939.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	12	5
AB255940.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	12	5
AB255945.1	Retained	Higher termite G	Japan Iriomote Island	[17]	37	5
AB255950.1	Retained	Higher termite G	Thailand Bangkok	[17]	41	5
EF453821.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF453822.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF453826.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF453857.1	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454021.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454057.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454284.2	Retained	Higher termite G	Costa Rica	[16]	37	5

EF454314.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454418.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454434.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454506.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454604.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454924.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454949.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF454981.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF455006.2	Retained	Higher termite G	Costa Rica	[16]	37	5
EF453831.2	Retained	Higher termite G	Costa Rica	[16]	56	5
EF454276.2	Retained	Higher termite G	Costa Rica	[16]	56	5
EF454303.2	Retained	Higher termite G	Costa Rica	[16]	56	5
EF454610.2	Retained	Higher termite G	Costa Rica	[16]	56	5
EF454823.2	Retained	Higher termite G	Costa Rica	[16]	56	5
EF454908.2	Retained	Higher termite G	Costa Rica	[16]	56	5
AB192078.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	14	6
AB192079.1	Retained	Higher termite G	Thailand Nakhon Pathom	[17]	14	6
AB192080.1	Retained	Higher termite G	Thailand Nakhon Pathom	[17]	14	6
AB192081.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	15	6
AB192082.1	Retained	Higher termite G	Thailand Prachinburi	[17]	15	6
AB192096.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	15	6
AB243276.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	15	6
AB248829.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	15	6
AB248830.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	15	6
AB255941.1	Retained	Higher termite G	Thailand Prachinburi	[17]	15	6
AB255942.1	Retained	Higher termite G	Thailand Prachinburi	[17]	15	6
AB192083.1	Retained	Higher termite G	Thailand Prachinburi	[17]	16	7
AB192084.1	Retained	Higher termite G	Thailand Prachinburi	[17]	16	7
AB192085.1	Retained	Higher termite G	Thailand Prachinburi	[17]	17	8
AB192086.1	Retained	Higher termite G	Thailand Nakhon Pathom	[17]	17	8
AB192088.1	Retained	Higher termite G	Thailand Prachinburi	[17]	17	8
AB192097.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	23	8
AB243275.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	17	8
AB192087.1	Retained	Higher termite G	Thailand Nakhon Pathom	[17]	18	9
AB192093.2	Retained	Higher termite G	Thailand Pathum Thani	[17]	26	10
AB192095.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	22	10
AB234547.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	24	11
AB243279.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	25	12
DQ676420.1	Retained	Freshwater sediment	France Orsay	[40]	27	13
AB255931.1	Retained	Higher termite G	Japan Iriomote Island	[17]	28	14
AB255933.1	Retained	Higher termite G	Japan Iriomote Island	[17]	30	14
AB255935.1	Retained	Higher termite G	Japan Iriomote Island	[17]	32	14
AB255937.1	Retained	Higher termite G	Japan Iriomote Island	[17]	34	14
AB255951.1	Retained	Higher termite G	Thailand Bangkok	[17]	42	14
EF454275.2	Retained	Higher termite G	Costa Rica	[16]	34	14
EF454318.2	Retained	Higher termite G	Costa Rica	[16]	34	14
EF454325.2	Retained	Higher termite G	Costa Rica	[16]	34	14
EF454459.2	Retained	Higher termite G	Costa Rica	[16]	34	14
EF454461.2	Retained	Higher termite G	Costa Rica	[16]	34	14

EF454475.2	Retained	Higher termite G	Costa Rica	[16]	34	14
EF454585.2	Retained	Higher termite G	Costa Rica	[16]	34	14
EF454628.2	Retained	Higher termite G	Costa Rica	[16]	34	14
EF454783.1	Retained	Higher termite G	Costa Rica	[16]	34	14
EF454888.2	Retained	Higher termite G	Costa Rica	[16]	34	14
AB255932.1	Retained	Higher termite G	Japan Iriomote Island	[17]	29	15
AB255943.1	Retained	Higher termite G	Japan Iriomote Island	[17]	35	15
AB255934.1	Retained	Higher termite G	Japan Iriomote Island	[17]	31	16
AB255936.1	Retained	Higher termite G	Japan Iriomote Island	[17]	33	16
AB255944.1	Retained	Higher termite G	Japan Iriomote Island	[17]	36	16
AB255946.1	Retained	Lower termite G	Japan Yakushima	[17]	38	17
AB255947.1	Retained	Lower termite G	Japan Okinawa	[17]	39	18
AB255948.1	Retained	Lower termite G	Japan Okinawa	[17]	40	19
AB255949.1	Retained	Lower termite G	Japan Amami Island	[17]	40	19
AB255952.1	Retained	Higher termite G	Thailand Bangkok	[17]	43	20
AB255953.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	44	21
AB255954.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	45	22
AB255955.1	Retained	Higher termite G	Japan Iriomote Island	[17]	46	23
AB255956.1	Retained	Higher termite G	Japan Iriomote Island	[17]	46	23
AB255957.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	47	24
AB255958.1	Retained	Higher termite G	Thailand Pathum Thani	[17]	48	25
EF453758.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453759.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453760.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF453761.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453764.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453765.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453768.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453769.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453771.2-EF453773.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453775.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453777.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF453778.2-EF453780.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453782.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453783.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453784.1	Retained	Higher termite G	Costa Rica	[16]	51	26
EF453785.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453786.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453787.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453789.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453790.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453792.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453793.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453796.1-EF453800.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453802.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453803.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453806.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453811.2	Retained	Higher termite G	Costa Rica	[16]	50	26

EF453812.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453814.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF453815.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453816.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453819.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453824.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453827.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453833.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453838.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453841.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF453849.2	Retained	Higher termite G	Costa Rica	[16]	53	26
EF453859.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453865.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453866.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453872.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453873.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453884.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453886.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453895.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453903.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453906.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453907.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453924.2-EF453926.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453929.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453931.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453932.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453935.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF453936.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453943.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453944.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453948.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453949.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453951.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453954.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453958.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453959.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453961.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453966.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453980.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453981.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453984.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453986.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453990.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454001.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454004.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454016.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454020.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454025.2-EF454027.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454039.2	Retained	Higher termite G	Costa Rica	[16]	50	26

EF454040.2-EF454042.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454052.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454056.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454058.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454059.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454061.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454064.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454069.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454070.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454074.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454079.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454086.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454087.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454089.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454092.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454097.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454098.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454100.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454101.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454105.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454107.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454109.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454117.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454118.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454124.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454125.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454127.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454128.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454135.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454136.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454140.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454148.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454150.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454151.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454153.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454162.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454165.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454166.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454168.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454171.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454172.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454173.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454174.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454180.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454186.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454190.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454192.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454193.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454195.2	Retained	Higher termite G	Costa Rica	[16]	50	26

EF454199.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454200.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454203.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454209.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454211.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454214.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454216.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454218.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454229.2-EF454233.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454236.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454241.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454243.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454244.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454249.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454252.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454254.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454256.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454260.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454261.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454265.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454266.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454270.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454274.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454278.2-EF454280.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454282.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454286.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454290.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454296.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454307.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454308.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454310.1-EF454313.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454315.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454316.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454319.1-EF454324.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454326.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454328.2-EF454339.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454341.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454342.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454344.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454345.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454347.2-EF454351.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454356.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454357.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454359.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454360.1	Retained	Higher termite G	Costa Rica	[16]	51	26

EF454361.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454364.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454365.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454367.2-EF454373.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454376.1	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454377.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454381.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454385.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454387.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454389.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454390.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454391.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454392.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454393.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454394.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454395.1-EF454398.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454399.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454400.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454401.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454404.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454407.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454408.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454410.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454412.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454413.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454415.2-EF454417.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454419.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454420.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454422.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454424.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454426.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454427.2-EF454429.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454431.1-EF454433.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454435.2-EF454441.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454443.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454445.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454446.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454448.2-EF454452.1	Retained	Higher termite G	Costa Rica	[16]	53	26
EF454454.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454456.1-EF454458.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454460.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454463.2-EF454465.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454467.2-	Retained	Higher termite G	Costa Rica	[16]	50	26

EF454471.2						
EF454473.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454474.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454476.2-						
EF454484.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454485.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454486.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454487.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454488.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454489.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454490.1	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454491.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454493.2-						
EF454496.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454499.2-						
EF454502.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454504.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454507.1	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454509.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454510.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454513.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454515.2-						
EF454519.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454520.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454521.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454522.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454524.1	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454526.2-						
EF454535.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454537.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454539.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454541.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454542.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454545.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454548.2-						
EF454553.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454554.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454555.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454556.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454558.1	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454559.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454560.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454561.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454563.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454565.2-						
EF454567.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454568.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454571.2-						
EF454578.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454580.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454587.2-						
EF454594.2	Retained	Higher termite G	Costa Rica	[16]	50	26

EF454597.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454598.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454601.2-EF454603.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454606.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454607.1-EF454615.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454617.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454618.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454620.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454622.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454623.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454625.1	Retained	Higher termite G	Costa Rica	[16]	53	26
EF454626.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454627.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454629.2-EF454631.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454634.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454636.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454639.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454641.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454642.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454644.2-EF454647.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454649.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454650.1-EF454652.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454654.2-EF454669.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454672.2-EF454682.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454685.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454686.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454688.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454689.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454690.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454692.1-EF454695.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454696.1	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454697.1	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454698.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454700.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454701.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454702.1	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454703.2-EF454707.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454710.2-EF454712.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454714.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454716.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454718.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454720.2	Retained	Higher termite G	Costa Rica	[16]	50	26

EF454724.2-EF454726.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454728.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454733.2-EF454739.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454741.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454742.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454746.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454747.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454752.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454753.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454755.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454757.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454761.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454763.2-EF454765.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454767.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454769.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454770.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454771.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454773.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454775.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454776.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454777.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454779.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454780.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454782.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454784.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454786.1-EF454788.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454795.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454796.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454803.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454806.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454810.1	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454813.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454818.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454821.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454827.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454833.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454839.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454841.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454843.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454853.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454856.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454858.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454861.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454866.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454873.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454878.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454880.1	Retained	Higher termite G	Costa Rica	[16]	51	26

EF454881.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454882.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454890.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454892.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454894.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454905.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454906.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454910.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454911.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454915.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454944.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454961.1	Retained	Higher termite G	Costa Rica	[16]	53	26
EF454962.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454964.2-EF454968.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454972.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454973.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454975.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454976.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454977.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454978.2	Retained	Higher termite G	Costa Rica	[16]	53	26
EF454980.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454985.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454986.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454990.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454991.2	Retained	Higher termite G	Costa Rica	[16]	51	26
EF454992.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454993.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF454994.1	Retained	Higher termite G	Costa Rica	[16]	53	26
EF454995.2-EF454998.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF455000.2-EF455009.2	Retained	Higher termite G	Costa Rica	[16]	50	26
EF453762.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF453766.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF453808.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF453985.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454104.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454121.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454152.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454258.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454352.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454355.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454366.1	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454442.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454543.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454611.1	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454619.1	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454715.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454745.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454778.2	Retained	Higher termite G	Costa Rica	[16]	52	27

EF454930.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454952.1	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454954.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454974.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454988.2	Retained	Higher termite G	Costa Rica	[16]	52	27
EF454298.2	Retained	Higher termite G	Costa Rica	[16]	54	28
EF454358.2	Retained	Higher termite G	Costa Rica	[16]	54	28
EF454384.2	Retained	Higher termite G	Costa Rica	[16]	54	28
EF454569.2	Retained	Higher termite G	Costa Rica	[16]	54	28
EF454722.2	Retained	Higher termite G	Costa Rica	[16]	54	28
EF454760.1	Retained	Higher termite G	Costa Rica	[16]	54	28
EF454789.2	Retained	Higher termite G	Costa Rica	[16]	54	28
AM982635.2	Retained	Pig F	France Rennes	U	55	29
EF190822.1-EF190825.1	Retained	Landfill	England NW	[12]	55	29
EF190827.1	Retained	Landfill	England NW	[12]	55	29
EF190829.1	Retained	Landfill	England NW	[12]	55	29
EF453861.2	Retained	Higher termite G	Costa Rica	[16]	57	30
EF454234.2	Retained	Higher termite G	Costa Rica	[16]	57	30
EF454354.2	Retained	Higher termite G	Costa Rica	[16]	57	30
EF454414.2	Retained	Higher termite G	Costa Rica	[16]	57	30
EF454514.2	Retained	Higher termite G	Costa Rica	[16]	57	30
EF454633.2	Retained	Higher termite G	Costa Rica	[16]	57	30
EF454762.2	Retained	Higher termite G	Costa Rica	[16]	57	30
EF454857.2	Retained	Higher termite G	Costa Rica	[16]	57	30
EF454970.2	Retained	Higher termite G	Costa Rica	[16]	57	30
EF454403.2	Retained	Higher termite G	Costa Rica	[16]	58	31
EF454859.2	Retained	Higher termite G	Costa Rica	[16]	59	32
EF520548.1	Retained	Acid-impacted lake	Adirondack lake NY	[41]	60	33
EF520549.1	Retained	Acid-impacted lake	Adirondack lake NY	[41]	60	33
EF651009.1	Retained	Soil	Australia 30.18 S 149.46 E	U	61	34
EU459511.1	Retained	capybara F	St. Louis Zoo. Park, MO	[19]	62	35
EU461471.1	Retained	black rhinoceros F	St. Louis Zoo. Park, MO	[19]	63	36
EU469557.1	Retained	Western lowland gorilla F	St. Louis Zoo. Park, MO	[19]	66	37
EU470332.1	Retained	Grevys zebra F	St. Louis Zoo. Park, MO	[19]	68	38
EU471816.1	Retained	Asiatic elephant F	St. Louis Zoo. Park, MO	[19]	69	39
EU473585.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	72	40
EU473589.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	73	41
EU771283.1	Retained	Asiatic elephant F	St. Louis Zoo. Park, MO	[19]	74	42
EU771291.1	Retained	Asiatic elephant F	St. Louis Zoo. Park, MO	[19]	75	43
EU771376.1	Retained	Asiatic elephant F	St. Louis Zoo. Park, MO	[19]	76	44
EU774390.1	Retained	Eastern black and white colobus F	St. Louis Zoo. Park, MO	[19]	77	45
EU774413.1	Retained	Eastern black and white colobus F	St. Louis Zoo. Park, MO	[19]	78	46
EU774455.1	Retained	Eastern black and white colobus F	St. Louis Zoo. Park, MO	[19]	79	47
EU774466.1	Retained	Eastern black and	St. Louis Zoo. Park, MO	[19]	80	48

		white colobus F				
EU779351.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	81	49
EU779354.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	82	50
EU779378.1	Retained	Somali wild ass F	St. Louis Zoo. Park, MO	[19]	83	51
GQ348358.1	Retained	Marine water	Canada British Columbia, Saanich Inlet	[14]	84	52
HQ163188.1	Retained	Marine water	Canada British Columbia, Saanich Inlet	U	86	53
EU101143.1	Retained	sulfidic biofilm	Italy Frasassi caves	[15]	87	54
EU101148.1	Retained	sulfidic biofilm	Italy Frasassi caves	[15]	87	54
EU101166.1	Retained	sulfidic biofilm	Italy Frasassi caves	[15]	87	54
EU101192.1	Retained	sulfidic biofilm	Italy Frasassi caves	[15]	87	54
FJ716839.1	Retained	Marine sediment	England Northumberland	U	88	55
AB252949.1	Gemmati-mondales	Iron-oxidaton biofilm	Japan Ishikawa, Kaga, Lagoon Shibayama	U		
AB385905.1	Gemmati-mondales	Ostrich cecum	Japan Mie	[42]		
AB385914.1	Gemmati-mondales	Ostrich cecum	Japan Mie	[42]		
AB385921.1	Gemmati-mondales	Ostrich cecum	Japan Mie	[42]		
AF234148.1	Gemmati-mondales	Soil	Australia Sturt National Park, New South Wales	[43]		
AB192082.1	Duplicate	Higher termite G	Thailand Prachinburi	[17]		
AB192088.1	Duplicate	Higher termite G	Thailand Prachinburi	[17]		
AB243275.1	Duplicate	Higher termite G	Thailand Pathum Thani	[17]		
AB243276.1	Duplicate	Higher termite G	Thailand Pathum Thani	[17]		
AB243277.1	Duplicate	Higher termite G	Thailand Pathum Thani	[17]		
AB255938.1	Duplicate	Higher termite G	Thailand Pathum Thani	[17]		
AB255940.1	Duplicate	Higher termite G	Thailand Pathum Thani	[17]		
AB386089.1	Duplicate	Ostrich cecum	Japan Mie	[42]		
AF224857.1	Duplicate	Marine water	Cariaco Basin	[44]		
EF454087.2	Duplicate	Higher termite G	Costa Rica	[16]		
EF454136.2	Duplicate	Higher termite G	Costa Rica	[16]		
EF454572.2	Duplicate	Higher termite G	Costa Rica	[16]		
EF454994.1	Duplicate	Higher termite G	Costa Rica	[16]		
EU461471.1	Duplicate	Black rhinoceros F	St. Louis Zoo. Park, MO	[19]		
EU463463.1	Duplicate	horse F	St. Louis Zoo. Park, MO	[19]		
GQ451318.1	Duplicate	Yunnan snub-nosed monkey F	China Tibet, Yunnan	[19]		
JQ346742.1	Duplicate	Bovine, R	*F. succinogenes* S85	U		
AF234139.1	Deferri-bacteries	Soil	Australia Sturt National Park, New South Wales	[43]		
AF234140.1	Deferri-bacteries	Soil	Australia Sturt National Park, New South Wales	[43]		
AB252948.1	Chlorobi	Iron-oxidaton biofilm	Japan Ishikawa, Kaga, Lagoon Shibayama	U		
AF234125.1	Chlorobi	Soil	Australia Sturt National Park, New South Wales	[43]		
AF406551.1	Chlorobi	Marine water	North Aegean	[45]		
EU775761.1	Chimera	Horse F	St. Louis Zoo. Park, MO	[19]		
EU779328.1	Chimera	Somali wild ass F	St. Louis Zoo. Park, MO	[19]		

AF406544.1	Acido-bacteria	Marine water	North Aegean	[45]
AM690809.1	Acido-bacteria	Freshwater	China Taihu Lake	[46]
AM690985.1	Acido-bacteria	Freshwater	China Taihu Lake	[46]
AY509521.1	Acido-bacteria	Freshwater	Sweden Lake Limmaren	[47]
AY571789.1	Acido-bacteria	Soil	Scott base, Antarctica	[48]
AY571790.1	Acido-bacteria	Soil	Scott base, Antarctica	[48]
AY571791.1	Acido-bacteria	Soil	Scott base, Antarctica	[48]
AY571792.1	Acido-bacteria	Soil	Scott base, Antarctica	[48]
AY571793.1	Acido-bacteria	Soil	Scott base, Antarctica	[48]
AY571794.1	Acido-bacteria	Soil	Scott base, Antarctica	[48]
AY571795.1	Acido-bacteria	Soil	Scott base, Antarctica	[48]
AY571796.1	Acido-bacteria	Soil	Scott base, Antarctica	[48]
AY571797.1	Acido-bacteria	Soil	Scott base, Antarctica	[48]
AY962277.1	Acido-bacteria	Freshwater	11 km south of Moscow, ID	[49]
DQ017910.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
DQ017915.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
DQ017918.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
DQ017924.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
DQ017927.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
DQ017934.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
DQ017939.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
DQ017940.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
DQ017945.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
DQ017946.1	Acido-bacteria	Freshwater	Schlitz 36110, Germany	[50]
FJ905643.1	Acido-bacteria	Undersea volcano	24°48.282′ S/177°1.141′ W	[51]
FJ905741.1	Acido-bacteria	Undersea volcano	21°9.205′ S/175°44.764′ W	[51]
AB385939.1	WS3	Ostrich cecum	Japan Mie	[42]
AB385982.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB385987.1	<900 bp	Ostrich cecum	Japan Mie	[42]

AB385992.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB385997.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386004.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386006.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386018.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386055.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386062.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386067.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386079.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386084.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386108.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386142.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB386172.1	<900 bp	Ostrich cecum	Japan Mie	[42]
AB476934.1-AB476952.1	<900 bp	Ostrich cecum	Japan Mie	[18]
AF165268.1	<900 bp	Soil	Kohala Forest Reserve, HA	[39]
AF165269.1	<900 bp	Soil	Kohala Forest Reserve, HA	[39]
AF165271.1	<900 bp	Soil	Kohala Forest Reserve, HA	[39]
AF165273.1	<900 bp	Soil	Kohala Forest Reserve, HA	[39]
AF224788.1	<900 bp	Marine water	Cariaco Basin	[44]
AF224822.1	<900 bp	Marine water	Cariaco Basin	[44]
AF224857.1	<900 bp	Marine water	Cariaco Basin	[44]
AF224872.1	<900 bp	Marine water	Cariaco Basin	[44]
AY095633.1-AY095635.1	<900 bp	Marine water	Northeast Pacific Ocean (32°50' N, 120°40' W)	U
AY095796.1	<900 bp	Marine water	Coral Sea (16°35.16' S, 166°14.82' E)	U
AY095797.1	<900 bp	Marine water	Coral Sea (16°35.16' S, 166°14.82' E)	U
AY095885.1	<900 bp	Marine water	Tahiti (17° S, 150° W)	U
AY145650.1	<900 bp	Hot spring	Mammoth Hot Springs	U
AY145655.1	<900 bp	Hot spring	Mammoth Hot Springs	U
AY311635.1	<900 bp	Yak, R	China Qilian Mountain	[52]
AY311689.1	<900 bp	Yak, R	China Qilian Mountain	[52]
AY509501.1	<900 bp	Freshwater	Sweden Lake Limmaren	[47]
AY509504.1	<900 bp	Freshwater	Sweden Lake Limmaren	[47]
AY509505.1	<900 bp	Freshwater	Sweden Lake Limmaren	[47]
AY509510.1	<900 bp	Freshwater	Sweden Lake Limmaren	[47]
AY571789.1	<900 bp	Soil	Kohala Forest Reserve, HA	[39]
AY869329.1-AY869346.1	<900 bp	Marine water	Greenland Sea, Arctic Ocean	[53]
AY869347.1-AY869380.1	<900 bp	Marine water	Ionian Sea, Eastern Mediterranean	[53]
DQ119083.1	<900 bp	Membrane bioreacter	Singapore	U
DQ501290.1	<900 bp	Freshwater	Germany Lake Breiter Luzin	U
DQ501365.1	<900 bp	Freshwater	Germany Lake Tiefwaren	U
DQ825779.1	<900 bp	Freshwater sediment	UK Gwynedd	U
EF072138.1	<900 bp	Soil	Watkinsville, GA (33540N, 83240W)	[54]
EF072208.1	<900 bp	Soil	Watkinsville, GA (33540N,	[54]

			83240W)	
EF072755.1	<900 bp	Soil	Watkinsville, GA (33540N, 83240W)	[54]
EF072758.1	<900 bp	Soil	Watkinsville, GA (33540N, 83240W)	[54]
EF074215.1	<900 bp	Soil	Watkinsville, GA (33540N, 83240W)	[54]
EF074226.1	<900 bp	Soil	Watkinsville, GA (33540N, 83240W)	[54]
EF074518.1	<900 bp	Soil	Watkinsville, GA (33540N, 83240W)	[54]
EF074547.1	<900 bp	Soil	Watkinsville, GA (33540N, 83240W)	[54]
EF075320.1	<900 bp	Soil	Watkinsville, GA (33540N, 83240W)	[54]
EF186234.1	<900 bp	Bovine, R	University of Illinois at Urbana-Champaign	[12]
EF186235.1	<900 bp	Bovine, R	University of Illinois at Urbana-Champaign	[12]
EF186236.1-EF186243.1	<900 bp	Landfill	England NW	[12]
EF186244.1-EF186248.1	<900 bp	Landfill	England Bidston Moss	[12]
EF186249.1-EF186267.1	<900 bp	Landfill	England NW	[12]
EF186268.1-EF186281.1	<900 bp	Landfill	England Bidston Moss	[12]
EF186282.1-EF186293.1	<900 bp	Landfill	England NW	[12]
EF520550.1	<900 bp	Acid-impacted lake	Adirondack Lake, NY	[47]
EF554981.1	<900 bp	Soil	France St. Paul lez Durance	U
EF651500.1	<900 bp	Soil	30.18 S 149.46 E	U
EF651576.1	<900 bp	Soil	30.18 S 149.46 E	U
EF662988.1	<900 bp	Soil	Michigen (42.40 N 85.40 W)	U
EF663737.1	<900 bp	Soil	Michigen (42.40 N 85.40 W)	U
EF664451.1	<900 bp	Soil	Michigen (42.40 N 85.40 W)	U
EF665397.1	<900 bp	Soil	Michigen (42.40 N 85.40 W)	U
EF665633.1	<900 bp	Soil	Michigen (42.40 N 85.40 W)	U
EF665686.1	<900 bp	Soil	Michigen (42.40 N 85.40 W)	U
EF681722.1	<900 bp	Bovine, non-R	Turkey Diyarbakir	U
EU297958.1	<900 bp	Soil	Kansas (39.0833N 96.5833W)	U
EU298010.1	<900 bp	Soil	Kansas (39.0833N 96.5833W)	U
EU298251.1	<900 bp	Soil	Kansas (39.0833N 96.5833W)	U
EU298944.1	<900 bp	Soil	Kansas (39.0833N 96.5833W)	U
EU299908.1	<900 bp	Soil	Kansas (39.0833N 96.5833W)	U
EU300162.1	<900 bp	Soil	Kansas (39.0833N 96.5833W)	U
EU300598.1	<900 bp	Soil	Kansas (39.0833N 96.5833W)	U
EU781744.1	<900 bp	Root tissue	Italy Caserta	[55]
EU981941.1	<900 bp	Buffalo, R	China Guangxi	U
FJ711708.1-FJ711715.1	<900 bp	Freshwater	England Priest Pot, Lake District, Cumbria	[13]
FJ711716.1-FJ711726.1	<900 bp	Freshwater sediment	England Priest Pot, Lake District, Cumbria	[13]

FJ711727.1-FJ711732.1	<900 bp	Freshwater	England Priest Pot, Lake District, Cumbria	[13]
FJ711733.1	<900 bp	Freshwater sediment	England Priest Pot, Lake District, Cumbria	[13]
FJ711734.1	<900 bp	Freshwater sediment	England Priest Pot, Lake District, Cumbria	[13]
FJ711735.1	<900 bp	Freshwater	England Priest Pot, Lake District, Cumbria	[13]
FJ711736.1-FJ711738.1	<900 bp	Freshwater sediment	England Priest Pot, Lake District, Cumbria	[13]
FJ711739.1-FJ711743.1	<900 bp	Soil	England Priest Pot, Lake District, Cumbria	[13]
FJ711744.1-FJ711746.1	<900 bp	Freshwater sediment	England Priest Pot, Lake District, Cumbria	[13]
FJ711747.1-FJ711750.1	<900 bp	Soil	England Priest Pot, Lake District, Cumbria	[13]
FJ711751.1	<900 bp	Sheep, non-R	England Priest Pot, Lake District, Cumbria	[13]
FJ711752.1	<900 bp	Sheep, non-R	England Priest Pot, Lake District, Cumbria	[13]
FJ711753.1	<900 bp	Soil	England Priest Pot, Lake District, Cumbria	[13]
FJ753112.1	<900 bp	Marine sediment	Canal St. Antoine, Gulf of Fos	U
FJ824887.1	<900 bp	marine sediment	Central North Sea (54°4′N/4°E)	[56]
FJ824888.1	<900 bp	marine sediment	Central North Sea (54°4′N/4°E)	[56]
FJ824897.1	<900 bp	marine sediment	Central North Sea (54°4′N/4°E)	[56]
FJ824900.1	<900 bp	marine sediment	Central North Sea (54°4′N/4°E)	[56]
FJ824901.1	<900 bp	marine sediment	Central North Sea (54°4′N/4°E)	[56]
FJ824906.1-FJ824908.1	<900 bp	marine sediment	Central North Sea (54°4′N/4°E)	[56]
GQ183237.1	<900 bp	Freshwater sediment	Pullman, WA	[57]
GQ505949.1	<900 bp	Bovine, R	Egypt Cairo	U
GU323642.1	<900 bp	Freshwater	China Nanjing	U
HM104720.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104722.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104731.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104735.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104754.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104756.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104767.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104806.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104816.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104820.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104821.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104828.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104868.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104911.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104958.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM104984.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM105466.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM105476.1	<900 bp	Bovine, R	Stillwater, OK	[57]
HM208520.1	<900 bp	Freshwater	China Nanjing	U

HQ031772.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ044682.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ044687.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ046088.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ047983.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ049319.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ050187.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ051872.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ068978.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ071706.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ073696.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ075324.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ076903.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ076962.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ079237.1	<900 bp	Oil sands	Alberta, Canada	[58]
HQ144020.1	<900 bp	Microbial mat (freshwater)	Canada Goodenough Lake	U
HQ386512.1	<900 bp	Freshwater	Sweden Lake Limmaren	U
JN412160.1	<900 bp	Biofilter substrate	France St. Pol de Leon	[59]
JN697125.1	<900 bp	Soil	India	U
JN865979.1	<900 bp	Freshwater	South Africa Vaal River, Scandinawiee	U
JQ815608.1	<900 bp	Acid-impacted water	Spain Tinto River Sediment	U
JQ906961.1	<900 bp	Swine anaerobic lagoon	Brazil	[60]
JQ937375.1	<900 bp	Freshwater	Israel Lake Kinneret	U
JQ937378.1	<900 bp	Freshwater	Israel Lake Kinneret	U
M62694.1	<900 bp	Sheep, R	*F. succinogenes* MM4	[37]

[a] R = Rumen contents, F = Feces, G = Gut

[b] Unpublished references are marked as "U"

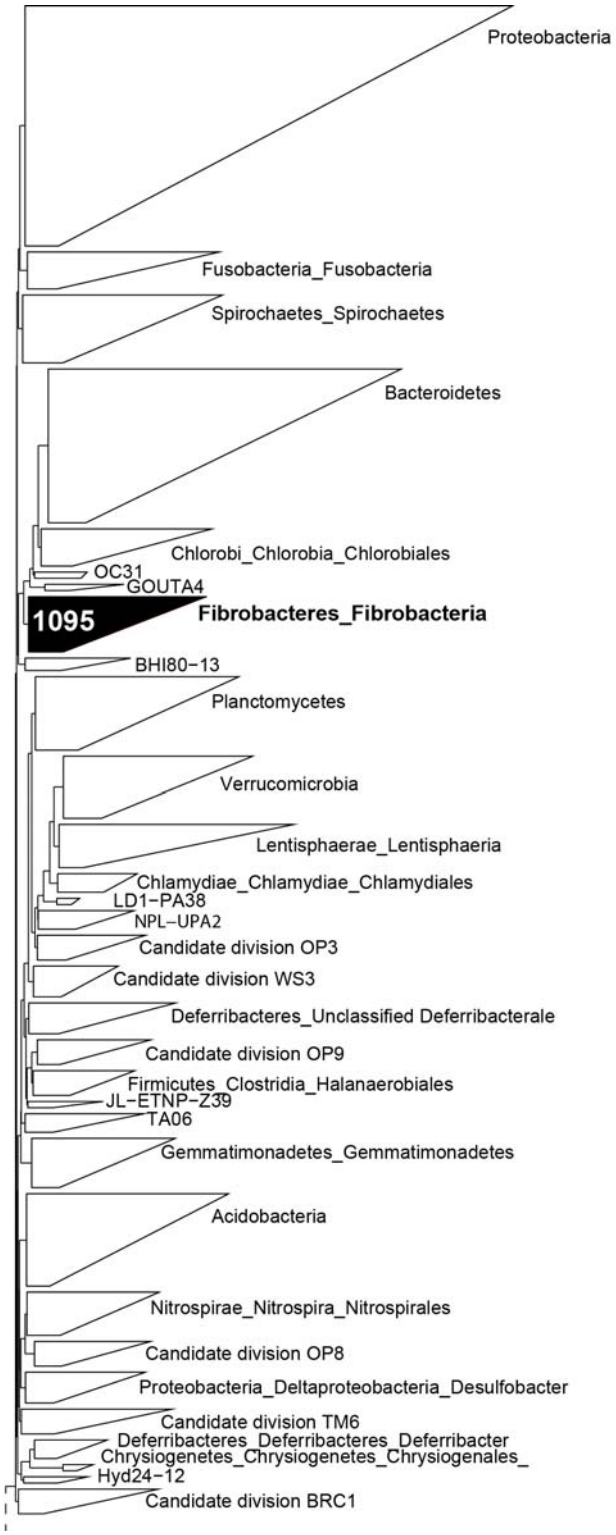

Figure S1. Cropped view of a phylogenetic tree for all *Fibrobacteres* 16S rRNA sequences generated using ARB and the Silva NR database. The *Fibrobacteres* have been emphasized with black fill, with the total number of sequences inside the *Fibrobacteres* clade are indicated.

An analysis of our sequences revealed that they were generated from both animal [4,8,9,10,16,17,19,33,34,35,37] and environmental [12,14,15,40,41] sources, including a wide range of locations spanning the globe (Fig. 1). This range of sampling locations likely reflects the distribution of research groups and interests rather than the dominant natural reservoirs for *Fibrobacteres* bacteria. The two largest groups of sequences belonged to either those associated with mammals (17 %, of which 9 % were specifically from ruminants) or termites (81 %), with the remaining sequences isolated from non host-associated environments (2 %) and a single isolate from the water flea *Ventiella sulfuris*. The rumen samples were nearly evenly split between sheep (40 sequences) and cows (32 sequences), with the remaining two sequences from goats. Among the non-ruminants there was a high degree of diversity (Fig. 1 and Table S1), with the most represented non-ruminant animal hosts being Somali wild asses (23 sequences), Yunnan snub-nosed monkeys (12 sequences), and the Eastern black and white colobus (7 sequences). Of the non-host associated environments, 3 were from marine sources, 7 from freshwater sources, 2 from soil, and the remaining 8 from landfills. Among the freshwater sources, surface water, acid-impacted lakes, and sulfidic cave waters were all represented.

Further analysis of the 245 *F. succinogenes* sequences only present in our short sequence database revealed that they were generated primarily from free-living environments, the bulk of which were from landfills (24 %), soil (16 %), and aquatic (38 %) sources. The 43 host-associated short sequences were dominated by rumen (10 %) and ostrich cecum (7 %) sources. Specific locations unique to the short sequence set included oil sands, yak and buffalo rumens, hot springs, ostrich cecae, and a swine anaerobic lagoon.

An initial *Fibrobacteres* phylogeny. To gain an initial understanding of the phylogenetic relationship between our identified *Fibrobacteres* sequences, we generated Neighbor-Joining

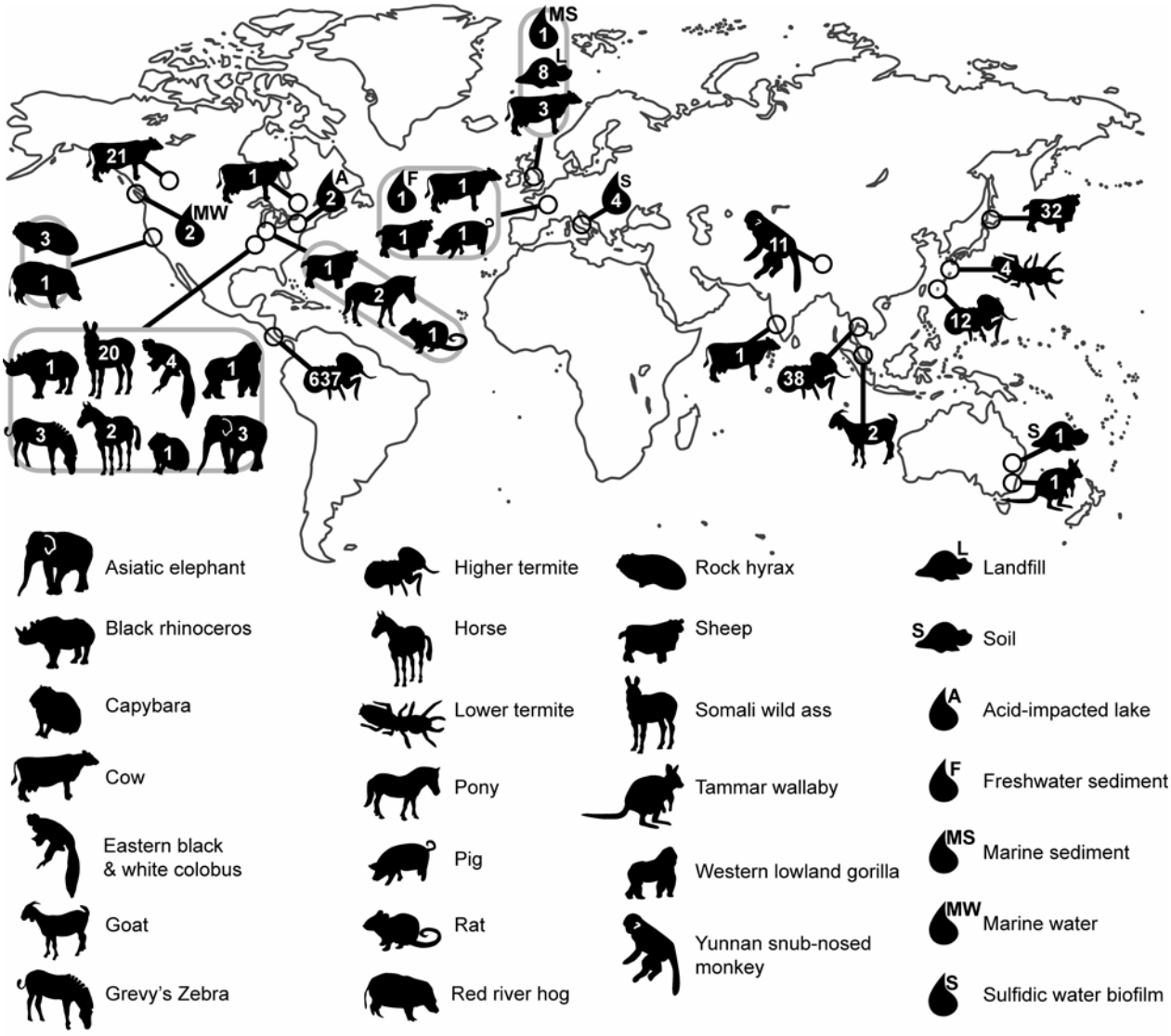

Figure 1. A world map showing the distribution and number of *Fibrobacteres* sequences used in this study. Sequence source is indicated by the silhouette, with the number of sequences per source given within the silhouette. Sample locations are indicated by the circles, and are based upon the reported locations in the relevant publication or GenBank accession information.

(NJ) trees using the combined long and short sequence databases (Fig. S2). In general, we found that *Fibrobacteres* sequences grouped according to their host or environmental association. For example, the higher and lower termites formed clades distinct from other host-associated sequences, while there was very little mixing of environmental (aquatic or soil) and host-associated sequences within terminal clades. Furthermore, we also found a single branch point that includes almost all of the host-associated sequences, with the *F. succinogenes* and *F. intestinalis* sequences forming distinct clades from that common branch point.

An OTU analysis of the phylum *Fibrobacteres*. A typical challenge in analyzing sequence libraries is the bias toward those samples that have the largest numbers of sequences. This is prevalent in our sequence database, as there is apparent numerical superiority of sequences from two sources: (a) the large number of historical studies focused on *Fibrobacteres* from ruminants, and (b) the inclusion of a particularly exhaustive study on the hindgut bacterial populations in higher termites [16]. One approach to reduce this complexity is to evaluate sequence libraries using operational taxonomic units (OTUs), which use percentages of sequence similarity to define phylogenetic relationships at different taxonomic levels, independent of sequence count. An OTU analysis of the long sequence database using mothur [25] revealed that there are at least 55 genera present within the phylum *Fibrobacteres* and at least 88 species (Table 1, OTUs at 97 % similarity for species and 95% for genus [61]). As an example of how OTU usage decreases the impact of sampling bias, of the 129 OTUs present at 97% sequence similarity in the short sequence database, those from higher termite guts were the source of slightly less than half of the OTUs (47.7%) despite encapsulating 80.6 % of all sequences. A single OTU for all 863 sequences was derived at 81 % similarity, approximately equal to the commonly used phylum-level cutoff of ≥ 80% similarity [61].

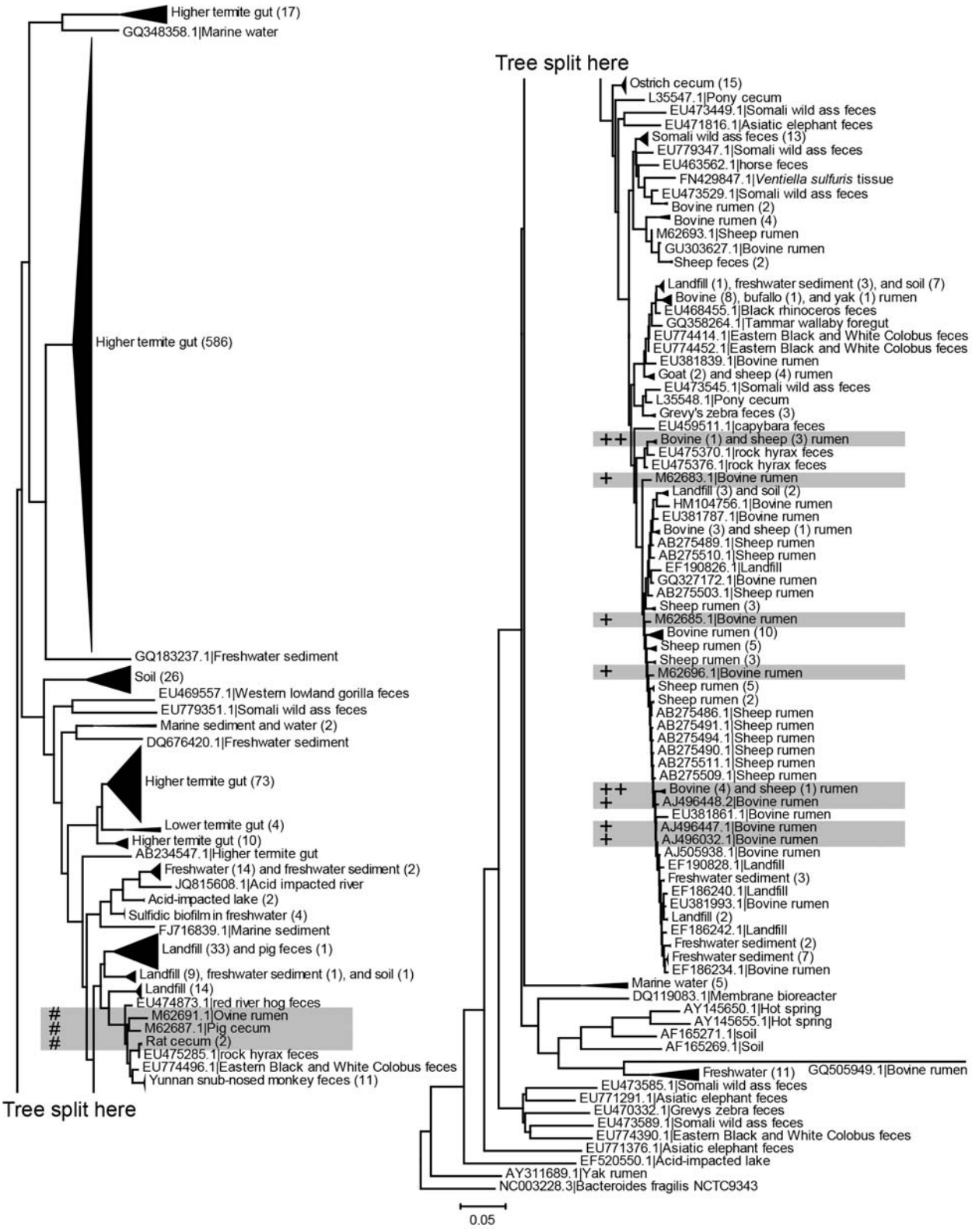

Figure S2. Neighbor-Joining tree of all Fibrobacteres sequences greater than 450 bp in length.

Fibrobacteres sequences of greater than 450 bp were used to used to construct this Neighbor-

Joining tree, with *Bacteroides fragilis* NCTC9343 as the outgroup. Where reasonable, clades have been collapsed to aid in visualization and the number of sequences in the clade indicated in parentheses; all branches with NCBI accession numbers represent single sequences. Sequences from cultured isolates have been highlighted in gray, with the locations of all *F. succinogenes* cultured strain sequences marked with (+), and of *F. intestinalis* with (#). The tree has been split as indicated by text for visual clarity.

A comparison of OTUs identified in the long and short sequence databases showed that, as expected, our short sequence database had a greater apparent diversity at each sequence similarity level examined down to 80 %, at which point all sequences converged into a single OTU (Table 1). The inclusion of additional environmental sequences in the short sequence database resulted in the formation of both unique OTUs and mixing with OTUs already defined from the long sequence database. We found the largest difference in OTU counts at 97 %, with an additional 41 OTUs identified using the short sequence database. A total of 12 of these OTUs were composed entirely of sequences unique to the short sequence database and each of these OTUs included at least two sequences. Two of these 12 OTUs were of animal origin (ostrich cecae and bovine rumen) while the remainder were of environmental origin (freshwater, marine water, soil, and oil sands); none of the OTUs had mixed origin (*e.g.*, no OTU containing both fresh and marine water sequences). By comparing the sequences within the OTUs we found that the inclusion of short sequences did not strengthen our phylogeny, except in suggesting that additional work should be done to generate longer *Fibrobacteres* sequences from environmental sources.

Fibrobacter is the only described genus in the phylum *Fibrobacteres* and our OTU analysis did not collapse the *F. intestinalis* and *F. succinogenes* sequences into a single genus at 95 % sequence similarity (Table 1), despite this being an accepted percent similarity used to define genera [61]. The *Fibrobacter* genus achieve resolution at a single OTU until 93 % similarity, which mirrors the results found previously of relatedness at 91-93 % sequence [8]. Without additional gene sequence information it is difficult to determine whether this relatively low percent identity is due to high diversity within the previously defined *Fibrobacter* genus or the 16S rRNA gene sequences used in this and previous studies.

Table 1. Number and approximate taxonomic level of *Fibrobacteres* OTUs by percentage of 16S rRNA gene sequence similarity.

	Long sequence database	Short sequence database	
% Similarity	**Number of OTUs**	**Number of OTUs**	**Taxonomic level**
97	88	129	
96	69	105	
95	55	89	Species[a]
94	40	76	
93	32	64	
92	29	53	
91	23	44	Genus[b]
90	22	36	
81	1	2	Phylum[c]
80	1	1	

[a] At 95 % similarity the *F. succinogenes*-identified long sequences formed a single OTU.

[b] At 91 % similarity the *Fibrobacter*-identified long sequences formed a single OTU.

[c] At 81 % similarity all Fibrobacteres-identified long sequences formed a single OTU.

An OTU-based Phylogeny of the *Fibrobacteres* and *Fibrobacter* spp. In order to create a comparative phylogeny we used representative sequences from each OTU at 97 % sequence similarity in our long sequence database to create a Bayesian phylogenetic tree, as shown in Fig. 2. Despite being considered the same species, the various strains of *F. succinogenes* and *F. intestinalis* failed to collapse into singular representative OTUs at this degree of sequence similarity. Of the 88 OTUs represented in Fig. 2, 27 included two or more sequences and 16 were composed entirely of termite-derived sequences. The remaining 11 OTUs were each dominated by specific sequence sources, such as ruminants; a full distribution is given in Fig. 3. As with the short sequence trees, the organization and topology of the sequences (Fig. S2) shows strong differences between environmental or insect- and mammal-derived *Fibrobacteres*.

To date, only 16 *F. succinogenes* cultured isolates have been reported and at least partially characterized: strains 128 [62], 095 [62], A3c [63], B1 [64], BL2 [64], F_E [4], GC5 [8], H [4], HM2 [8], MB4 [8], MC1 [8], MM4 [8], R [4], REH9-1 [65], the type strain S85 [66], and U [4]. These *F. succinogenes* 16S rRNA sequences were classified into five species-level OTUs using a 97% identity for species (Table 1). The largest of these OTUs contained ruminal strains A3C, B1, BL2, S85 (both GenBank M62696 and AJ496566 accessions), and U all from cattle rumens; HM2 from a sheep rumen; and 56 other sequences of which 52 were ruminal and two each were from rock hyrax feces or landfills. Both REH9-1 (cattle rumen) and MB4 (sheep rumen) were single-sequence OTUs. GC5 (cattle feces) with 11 other sequences formed a diverse 97% OTU with origins from cattle, sheep, and goat rumens (7 sequences total), Eastern black and white colobus feces (2 sequences), black rhinoceros feces (1 sequence), and the Tammar wallaby foregut (1 sequence). MC1 (sheep rumen) formed an OTU with a single other sequence from a cattle rumen.

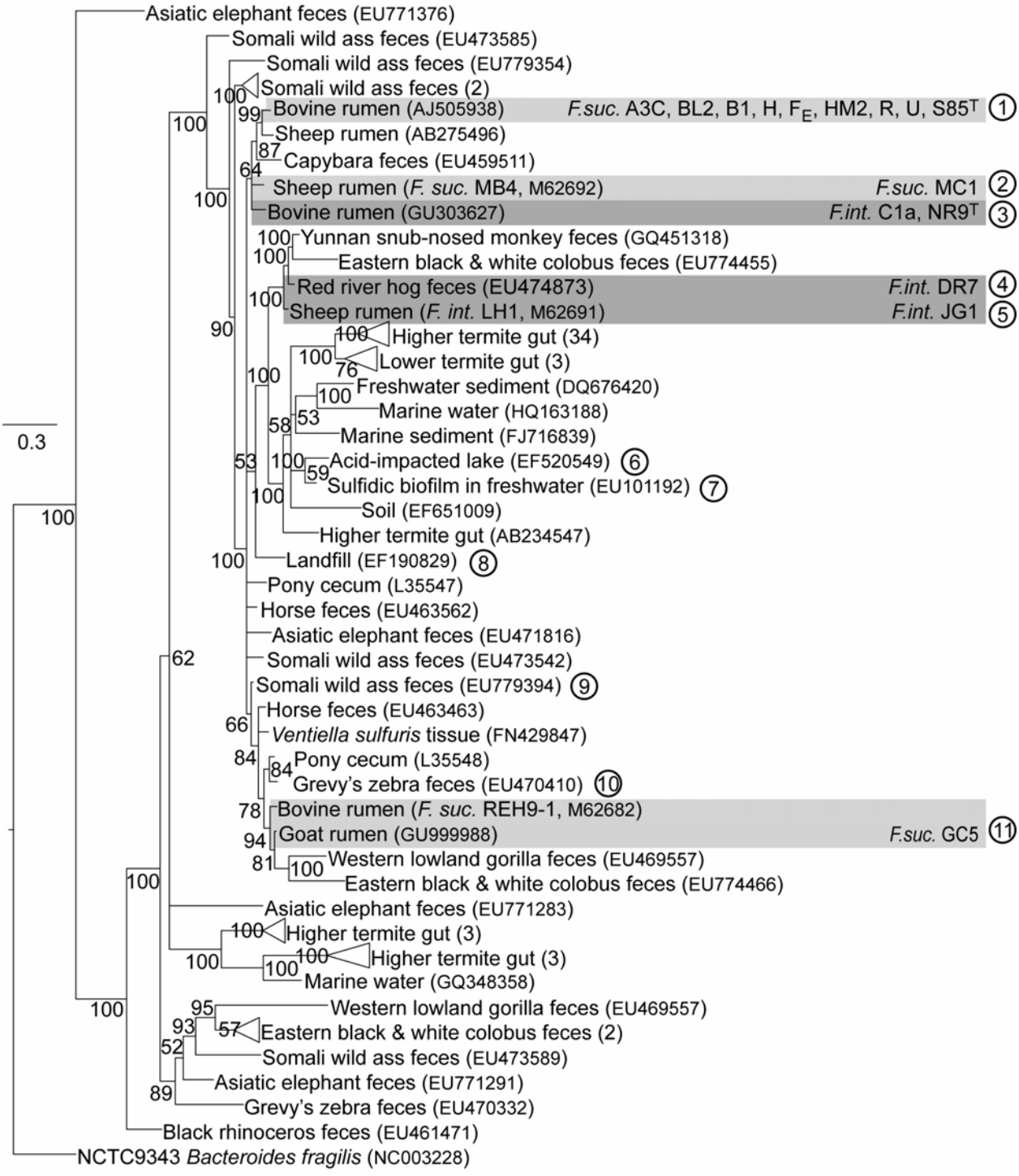

Figure 2. A Bayesian phylogenetic reconstruction of the representative 16S rRNA OTUs at 97% similarity with *Bacteroides fragilis* NCTC9343 as an outgroup. All posterior probability values are shown (ngen=10,000,000). Large clades of a single sample source type (*e.g.*, rumen)

have been collapsed, with the number of representative sequences indicated in parentheses. GenBank accession numbers for all singlets are indicated in parenthesis. Representative sequences including *F. succinogenes* are highlighted in light gray, while those including *F. intestinalis* are highlighted in dark gray. *F. succinogenes* = *F.suc.*; *F. intestinalis* = *F.int*. Arbitrary OTU numbers (in empty circles) have been assigned to correspond with Fig. 3.

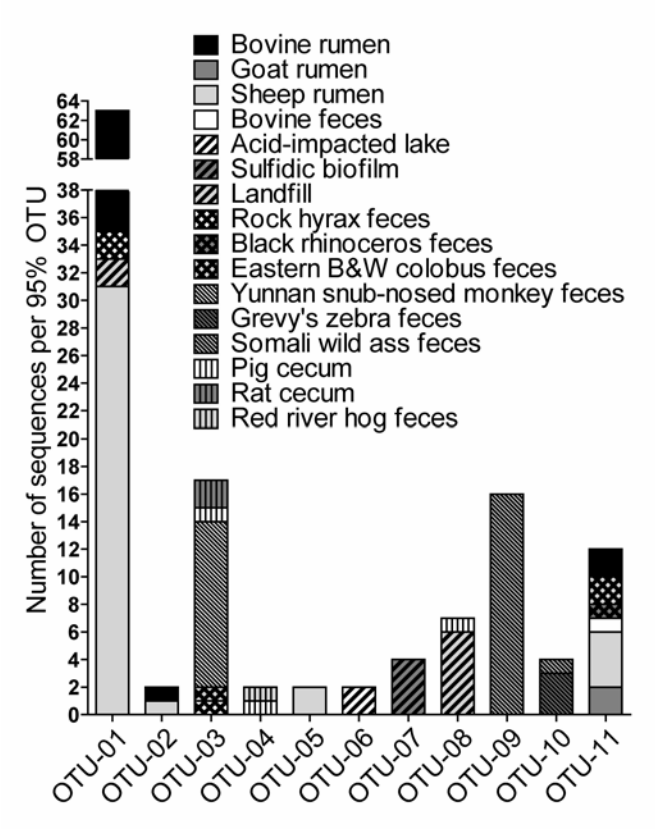

Figure 3. Distribution of sequence sources in non-termite 95% OTUs with two or more sequences. Arbitrary OTU numbers have been used, and correspond to those given in Fig. 2 (in empty circles).

For *F. intestinalis*, 16S rRNA sequences are available for all five published isolates including strains C1a [67] and DR7 [8] from porcine ceca; JG1 [8] and LH1 [8] from sheep ceca; and the type strain NR9 [65] from a rat cecum. At 97 % sequence similarity we found three separate *F. intestinalis*-containing OTUs. JGI and LH1 (sheep rumen) formed one OTU, while DR7 and a single sequence from red river hog feces formed a second. The third and largest OTU contained NR9 (rat cecum) and C1a (pig cecum) along with 15 other sequences: 12 from black and white colobus feces and singlets from a rat cecum, rock hyrax feces, and an eastern black and white colobus. All *F. intestinalis* sequences formed a single OTU at 95 % similarity, separate from the *F. succinogenes* OTU.

Higher and lower termite OTUs. In addition to the mammal-associated *Fibrobacter* species, numerous 16S rRNA sequences correlated to the *Fibrobacteres* have been reported from termites. These include the lower (wood-consuming) and higher (fungus-farming or detritivoric) termites [68]. Our OTU analysis agreed with previous work [17] in that the four lower termite-originating sequences do not combine with any higher termite sequences, at least not until the 90% OTU cutoff. The higher termite *Fibrobacteres* sequences formed multiple lineages (Fig. 2), and both lower and higher termite representative clades were intermixed with aquatic and soil-derived sequences. Even at the 90 % OTU there was no mixing of termite with non-termite sequences within an OTU.

Discussion

Members of the phylum *Fibrobacteres* are typified by two species, *F. succinogenes* and *F. intestinalis*, which formulate the current assumption that all members of this phylum are host-associated, non-motile, obligate anaerobes [7]. Scientific interest in the *Fibrobacteres* is

centered primarily on *Fibrobacter succinogenes* S85, which is a prolific and efficient microbe capable of degrading plant cell wall polysaccharides [4,6,69]. Given the importance of this species, it is probable that there are other members of the *Fibrobacteres* that would also be of interest in terms of biomass processing and cellulose degradation. Our broad phylogenetic analysis of this phylum is a first step towards understanding their diversity, environmental associations, and geographic distribution. Specifically, we have presented a global analysis of the geographic and phylogenetic distribution of the *Fibrobacteres* by generating an extensive phylogeny using all publicly available 16S rRNA gene sequences of 900 bp or longer. Shorter sequences were not included in our primary phylogeny due to the difficulties associated with separating spurious from actual diversity when mixing shorter-read data sets generated from non-overlapping variable regions of 16s rRNA sequence [20,70], but were used to construct a second, less stringent phylogeny for comparative purposes.

Sequences classified as belonging to the *Fibrobacteres* are present in widely disparate host-associated and free-living environments, indicating that the distribution and diversity within the *Fibrobacteres* is greater than previously thought. Our analysis revealed several distinct clades corresponding to both environmental or host associations, including a clear separation of mammal, insect and environment-associated sequences. Based on our pylogeny and recent phylogenetic and microscopic work [12,13,17], it is clear that our definition of the phylum *Fibrobacteres* must be re-evaluated. The current classification of *F. succinogenes* as a mammalian gut-specific microbe, for example, does not encompass those sequences that have been recovered from non-host associated environments (such as [12,13,14,15,39,40]). These data underscore the need for future work, specifically in isolating and characterizing non-host-associated cultures of *Fibrobacteres*, to refine our definition and understanding of this phylum.

In addition, we found that, based on commonly accepted levels of sequence similarity [61], the sequences currently defined as *F. succinogenes* and *F. intestinalis* do not represent single species or combine as a genus until highly relaxed similarity values are applied. In particular, the published 16S rRNA sequences from isolates of different *F. succinogenes* strains did not become a single OTU until 95 % sequence similarity, below the 97 % typically used to define a species [61]. Previous work by Amann *et al*. and Shinkai *et al*. using 16S rRNA gene sequencing showed that strains of *F. succinogenes* can be divided into four groups [8,33], with *F. intestinalis* strains forming two groups [8]. In our work, at 97 % sequence similarity, the *F. succinogenes* laboratory isolate sequences fell into five groups. Our phylogeny showed similar patterns for *F. succinogenes*, but we found only three major groups (Fig. 2). In particular, *F. succinogenes* MM4, MB4, and HM2, which were shown previously as a separate group [8,33] were found intermixed with other named sequences in our results. The separation of named strains shown in Fig. 2 for *F. intestinalis* were identical to the previously reported clusters [8]. It is probable that the sequence diversity within both "species" represents at least sub-species difference, as has been previously suggested [8] and codified for *F. succinogenes* subsp. *succinogenes* (type strain S85) and *F. succinogenes* subsp. *elongatus* (type strain HM2) [7].

It is intriguing to note that the phylogenetic division between the higher and lower termites is recapitulated in the division of their gut *Fibrobacteres* sequences, as was suggested by previous work done on the phylogeny of insect-associated *Fibrobacter* sequences [17]. It is possible that this separation is due to differences in the diets or physiology of the termites (lower termites consume wood, while higher termites cultivate fungi [68]), or that there are specific host-associated factors involved in a *Fibrobacteres*-termite symbiosis. Further work in the

termites is needed to determine if there exist distinct phylogenetic lineages of *Fibrobacteres* that are linked to host evolution, as has been found for other insect symbioses [71,72].

In conclusion, we have presented an extensive phylogenetic view of the phylum *Fibrobacteres* created from a synthesis of sequences reported from a wide range of scientific studies representing disparate sample sources and locations. With this framework it is now possible to redefine our understanding of this phylum. For example, the presence of verified *Fibrobacteres* sequences in oxygenated waters [15] strongly suggests that the current definition of this phylum containing only strict anaerobes may be inaccurate. It is also conceivable, given the wide host and environmental distribution reported on here, that additional *Fibrobacteres* isolates may be found that are more amenable to potential industrial exploitation than *F. succinogenes* [73,74,75] (*e.g.*, by aerotolerance or less-stringent nutritional requirements). The phylogenetic work presented here will help us better understand the diversity and potential environmental niches for these bacteria.

Acknowledgements

This work was funded by the Wisconsin Bioenergy Initiative supporting S.M.A and G.S. J.J.S. was supported by a University of Wisconsin Distinguished Fellowship. We would like to thank P. Weimer, F. Aylward, K. Grubbs, and A. Hanshew for useful comments and discussion.

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. Nature Reviews Microbiology 6: 121-131.
2. van Houtert MFJ (1993) The production and metabolism of volatile fatty acids by ruminants fed roughages: A review. Animal Feed Science and Technology 43: 189-225.
3. Hungate R (1950) The anaerobic mesophilic cellulolytic bacteria. Bacteriology Reviews 14: 1-49.
4. Bera-Maillet C, Ribot Y, Forano E (2004) Fiber-degrading systems of different strains of the genus *Fibrobacter*. Applied and Environmental Microbiology 70: 2172-2179.
5. Weimer PJ, French A, Calamari Jr. T (1991) Differential fermentation of cellulose allomorphs by ruminal cellulolytic bacteria. Applied and Environmental Microbiology 57: 3101-3106.
6. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, et al. (2011) The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. PLOS ONE.
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 Bacteriology 38: 430-435.
8. Amann R, Li C, Key R, Montgomery L, Stahl DA (1992) Diversity among *Fibrobacter* isolates: towards a phylogenetic classification. Systematic and Applied Microbiology 15: 23-31.
9. Lin C, Stahl DA (1995) Taxon-specific probes for the cellulolytic genus *Fibrobacter* reveal abundant and novel equine-associated populations. Applied and Environmental Microbiology 61: 1348-1351.
10. Whitford MF, Forster RJ, Beard CE, Gong J, Teather RM (1998) Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA Genes$^{\beta}$. Anaerobe 4: 153-163.
11. Kim M, Morrison M, Yu Z (2011) Status of the phylogenetic diversity census of ruminal microbiomes. FEMS Microbiology Ecology 76: 49-63.
12. 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. Environmental Microbiology 10: 1310-1319.
13. McDonald JE, de Menezes AB, Allison HE, McCarthy AJ (2009) Molecular biological detection and quantification of novel *Fibrobacter* populations in freshwater lakes. Applied and Environmental Microbiology 75: 5148-5152.
14. Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, et al. (2009) Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. Science 326: 578-582.
15. Macalady JL, Dattagupta S, Schaperdoth I, Jones DS, Druschel GK, et al. (2008) Niche differentiation among sulfur-oxidizing bacterial populations in cave waters. ISME Journal 2: 590-601.

16. Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450: 560-565.
17. Hongoh Y, Deevong P, Hattori S, Inoue T, Noda S, et al. (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. Applied and Environmental Microbiology 72: 6780-6788.
18. Matsui H, Ban-Tokuda T, Wakita M (2010) Detection of fiber-digesting bacteria in the ceca of ostrich using specific primer sets. Current Microbiology 60: 112-116.
19. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, et al. (2008) Evolution of mammals and their gut microbes. Science 320: 1647-1651.
20. Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, et al. (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME Journal 4: 642-647.
21. DeSantis T, Hugenholtz P, Larsen N, Rojas M, Brodie E, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench campatible with ARB. Applied and Environmental Microbiology 72: 5069-5072.
22. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37: D141-D145.
23. Pruesse E, Quast C, Knittle K, Fuches B, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35: 7188-7196.
24. Ludwig W, Strunk O, Westram R, Richter L, Meier H, et al. (2004) ARB: a software environment for sequence data. Nucleic Acids Research 32: 1363-1371.
25. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing MOTHUR: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75: 7537-7541.
26. Huelsenbeck J, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics Applications Note 17: 754-755.
27. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.
28. Rambaut A (2009) FigTree v1.3.1.
29. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.
30. Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences of the United States of America 101: 11030-11035.
31. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution Submitted.
32. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23: 127-128.

33. Shinkai T, Ohji R, Matsumoto N, Kobayashi Y (2009) Fibrolytic capabilities of ruminal bacterium *Fibrobacter succinogenes* in relation to its phylogenetic grouping. FEMS Microbiology Letters 294: 183-190.
34. Monteils V, Cauquil L, Combes S, Godon J-J, Gidenne T (2008) Potential core species and satellite species in the bacterial community within the rabbit caecum. FEMS Microbiology Ecology 66: 620-629.
35. Kong Y, Teather R, Forster R (2010) Composition, spatial distribution, and diversity of the bacterial communities in the rumen of cows fed different forages. FEMS Microbiology Ecology 74: 612-622.
36. Pope PB, Denman SE, Jones M, Tringe SG, Barry K, et al. (2010) Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. Proceedings of the National Academy of Sciences.
37. Amann R, Krumholz L, Stahl DA (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. Journal of Bacteriology 172: 762-770.
38. Wu J, Yan F, Gao R, Huang Z, Xu B, et al. (2010) Study of fecal bacterial diversity in Yunnan snub-nosed monkey (*Rhinopithecus bieti*) using phylogenetic analysis of cloned 16S rRNA gene sequences. African Journal of Biotechnology 9: 6278-6289.
39. Nüsslein K, Tiedje J (1999) Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. Applied and Environmental Microbiology 65: 3622-3626.
40. Briée C, Moreira D, López-García P (2007) Archaeal and bacterial community composition of sediment and plankton from a suboxic freshwater pond. Research in Microbiology 158: 213-227.
41. Percent SF, Frischer ME, Vescio PA, Duffy EB, Milano V, et al. (2008) Bacterial community structure of acid-impacted lakes: what controls diversity? Applied and Environmental Microbiology 74: 1856-1868.
42. Matsui H, Kato Y, Chikaraishi T, Moritani M, Ban-Tokuda T, et al. (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.
43. Holmes AJ, Bowyer J, Holley MP, O'Donoghue M, Montgomery M, et al. (2000) Diverse, yet-to-be-cultured members of the *Rubrobacter* subdivision of the Actinobacteria are widespread in Australian arid soils. FEMS Microbiology Ecology 33: 111-120.
44. Madrid VM, Taylor GT, Scranton MI, Chistoserdov AY (2001) Phylogenetic diversity of bacterial and archaeal communities in the anoxic zone of the Cariaco Basin. Applied and Environmental Microbiology 67: 1663-1674.
45. Moeseneder MM, Arrieta JM, Herndl GJ (2005) A comparison of DNA- and RNA-based clone libraries from the same marine bacterioplankton community. FEMS Microbiology Ecology 51: 341-352.
46. Wu QL, Zwart G, Wu J, Kamst-van Agterveld MP, Liu S, et al. (2007) Submersed macrophytes play a key role in structuring bacterioplankton community composition in the large, shallow, subtropical Taihu Lake, China. Environmental Microbiology 9: 2765-2774.
47. Eiler A, Bertilsson S (2004) Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. Environmental Microbiology 6: 1228-1243.

48. Saul DJ, Aislabie JM, Brown CE, Harris L, Foght JM (2005) Hydrocarbon contamination changes the bacterial diversity of soil from around Scott Base, Antarctica. FEMS Microbiology Ecology 53: 141-155.
49. Ball C, Crawford R (2006) Bacterial diversity within the planktonic community of an artesian water supply. Canadian Journal of Microbiology 52: 246-259.
50. Beier S, Witzel K-P, Marxsen J (2008) Bacterial community composition in central European running waters examined by temperature gradient gel electrophoresis and sequence analysis of 16S rRNA genes. Applied and Environmental Microbiology 74: 188-199.
51. Forget NL, Murdock SA, Juniper SK (2010) Bacterial diversity in Fe-rich hydrothermal sediments at two South Tonga Arc submarine volcanoes. Geobiology 8: 417-432.
52. An D, Dong X, Dong Z (2005) Prokaryote diversity in the rumen of yak (*Bos grunniens*) and Jinnan cattle (*Bos taurus)* estimated by 16S rDNA homology analyses. Anaerobe 11: 207-215.
53. Zaballos M, López-López A, Ovreas L, Bartual SG, D'Auria G, et al. (2006) Comparison of prokaryotic diversity at offshore oceanic locations reveals a different microbiota in the Mediterranean Sea. FEMS Microbiology Ecology 56: 389-405.
54. Jangid K, Williams MA, Franzluebbers AJ, Sanderlin JS, Reeves JH, et al. (2008) Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. Soil Biology and Biochemistry 40: 2843-2853.
55. Del Giudice L, Massardo D, Pontieri P, Bertea C, Mombello D, et al. (2008) The microbial community of Vetiver root and its involvement into essential oil biogenesis. Environmental Microbiology 10: 2824-2841.
56. Sapp M, Parker ER, Teal LR, Schratzberger M (2010) Advancing the understanding of biogeography–diversity relationships of benthic microorganisms in the North Sea. FEMS Microbiology Ecology 74: 410-429.
57. Fernando S, Purvis II H, Najar F, Sukiharnikov L, Kehbiel C, et al. (2010) Rumen microbial population dynamics during adaptation to a high-grain diet. Applied and Environmental Microbiology 76: 7482-7490.
58. Ramos-Padrón E, Bordenave S, Lin S, Bhaskar I, Dong X, et al. (2011) Carbon and sulfur cycling by microbial communities in a gypsum-treated oil sands tailings pond. Environmental Science & Technology 45: 439-446.
59. Renault D, Vallance J, Déniel F, Wery N, Godon J, et al. (2012) Diversity of bacterial communities that colonize the filter units used for controlling plant pathogens in soilless cultures. Microbial Ecology 63: 170-187.
60. Cardinali-Rezende J, Pereira Z, Sanz J, Chartone-Souza E, Nascimento A (2012) Bacterial and archaeal phylogenetic diversity associated with swine sludge from an anaerobic treatment lagoon. World Journal of Microbiology and Biotechnology 28: 3187-3195.
61. Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Applied and Environmental Microbiology 71: 1501-1506.
62. Gaudet G, Cheng KJ (1990) Use of glucose and cellobiose by 3 strains of *Fibrobacter succinogenes*. Reproduction Nutrition Development Supp 2: 201s-202s.
63. Dehority BA (1963) Isolations and characterization of several cellulolytic bacteria from *in vitro* rumen fermentations. Journal of Dairy Science 46: 217-222.

64. Stewart C, Paniagua C, Dinsdale D, Cheng K-J, Garrow S (1981) Selective isolation and characteristics of *Bacteroides succinogenes* from the rumen of a cow. Applied and Environmental Microbiology 41: 504-510.
65. Montgomery L, Macy J (1982) Characterization of rat cecum cellulolytic bacteria. Applied and Environmental Microbiology 44: 1435-1443.
66. Bryant M, Robinson I, Chu H (1959) Observations on the nutrition of *Bacteroides succinogenes* - a ruminal cellulolytic bacterium. Journal of Dairy Science 42: 1831-1847.
67. Varel V, Fryda S, Robinson I (1984) Cellulolytic bacteria from pig large intestine. Applied and Environmental Microbiology 47: 219-221.
68. Brune A, Friedrich M (2000) Microecology of the termite gut: structure and function on a microscale. Current Opinion in Microbiology 3: 263-269.
69. Maglione G, Russell JB, Wilson D (1997) Kinetics of cellulose digestion by *Fibrobacter succinogenes* S85. Applied and Environmental Microbiology 63: 665-669.
70. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environmental Microbiology 12: 118-123.
71. Aanen DK, Eggleton P, Rouland-LefÃ¨vre C, Guldberg-FrÃ¸slev T, Rosendahl Sr, et al. (2002) The evolution of fungus-growing termites and their mutualistic fungal symbionts. Proceedings of the National Academy of Sciences 99: 14887-14892.
72. Ohkuma M, Noda S, Hongoh Y, Kudo T (2001) Coevolution of symbiotic systems of termites and their gut microorganisms. RIKEN Review 41: 73-74.
73. Wen T-N, Chen J-L, Lee S-H, Yang N-S, Shyur L-F (2005) A truncated *Fibrobacter succinogenes* 1,3-1,4-β-D-Glucanase with improved enzymatic activity and thermotolerance. Biochemistry 44: 9197-9205.
74. Chen B-Y, Wang H-T (2008) Utility of enzymes from *Fibrobacter succinogenes* and *Prevotella ruminicola* as detergent additives. Journal of Industrial Microbiology & Biotechnology 35: 923-930.
75. Zeikus JG, Jain MK, Elankovan P (1999) Biotechnology of succinic acid production and markets for derived industrial products. Applied Microbiology and Biotechnology 51: 545-552.

APPENDIX TWO

16S rRNA-based community analyses are biased by initial PCR amplification

Jewell*, Kelsea A., Elizabeth Strassman, Caroline A. McCormick, Alissa S. Hanshew, Heidi Goodrich-Blair, Katherine D. McMahon, Nasia Safdar, Cameron R. Currie, Garret Suen*.

*Co-corresponding authors

KJ, ES, CM, and AH collected and processed samples, and performed sequencing reactions. HG-B, KM, NS, and CC supplied samples. KJ wrote the manuscript. KJ and CM performed sequence analyses. ES, CM, AH, and GS assisted with writing the manuscript. KJ and GS designed the experiments.

This manuscript is being re-formatted and edited for submission to the *Journal of Microbiological Methods*.

Abstract

Library generation can significantly impact sequence generation in large-scale bacterial community 16S rRNA sequencing studies. We examined the bias of library amplification on the detected bacterial community in four environmentally-derived samples. Genomic DNA was extracted from four sources (human feces, lake water, a whole firefly, and cow rumen solids) using template-specific, standard protocols. The 16S rRNA gene V5-V8 region was amplified using a high fidelity (PfuTurbo) and two low fidelity (EconoTaq and Platinum Blue PCR Supermix) polymerases, and one polymerase with slowed processivity and increased DNA binding affinity (SlowTaq); sequenced using 454 pyrosequencing, and analyzed using the program mothur. Three technical replicates were performed for the firefly and human fecal samples. Quantitative PCR (qPCR) was used to determine the concentration of bacterial DNA in each sample. Low-fidelity polymerases required the most modification of PCR conditions. Sequence sets tightly clustered by sample (principal component analysis and non-parametric multidimensional scaling), with minimal differences (Morisita-Horn similarity coefficient) within template type for all but firefly. Chimera formation was inconsistent. The percent of identical genus-level operation taxonomic units (OTUs) within each template was low (9.4 - 22.7 %), but was improved by removing low-abundance OTUs. Technical variation was minimal for the human fecal sample and high for the firefly (Θ_{YC} distances). By qPCR the proportion of bacterial DNA was 47.0 - 6.3 % for all but firefly (0.002 %). High fidelity polymerase usage did not decrease chimera formation but did increase the ease of making amplicon pools. Polymerase fidelity did not correlate with sequence set composition. Having a high input of bacterial DNA should decrease PCR-generated bias. OTUs with ≤ 1 sequence in replicated samples, and phyla with < 0.3 % relative sequence abundance, should not be included in multi-study comparisons.

Introduction

The current technological standard for non-culture based microbial ecology research is high-throughput, large-scale sequencing. Such sequencing methods can use pure or mixed DNA templates, allowing for the relatively rapid and inexpensive identification of microbes in complex communities [1], functional metagenomics [2], whole genome sequencing [3] and genotyping of single nucleotide polymorphisms [4]. In all of these applications, there is an existing and growing concern regarding the repeatability and reliability of the data generated, especially as sequencing results can significantly influenced by sample collection and extraction methods [5,6,7,8], and DNA amplification and processing [9,10,11,12,13,14]. The importance of each of these factors on the evaluation of the final sequence results is often not considered, and it is extremely difficult to separate technique-dependent variation from biologically important sequence differences.

In a large-scale evaluation of DNA polymerase enzymes, it was shown that polymerase selection for the generation of a sequencing pool has a major impact on the coverage and fidelity of whole-genome sequencing from pure templates [9]. Also, library preparation techniques (such as gel extraction and column clean-up) can shift both sequence length and the GC content of the final amplicon pool [10]. Moreover, library preparation location, such as between sequencing centers or laboratories, can have a measurable impact on the final detectable set of sequences [11] although this has been disputed [8]. When other variables are kept constant, switching DNA polymerases mid-way during a sequencing study can have an even greater impact [13] than changing sequencing platforms [15], although this has not been investigated for multiple polymerases. Taken together, these studies emphasize concerns that broad-scale comparisons of next-generation sequence databases are highly biased towards Type I errors

(false significance) resulting from compounded location, platform, extraction method, and variability in library creation.

Here, we surveyed a variety of DNA polymerase enzymes for their ability to consistently amplify the 16S rRNA gene (a common sequencing target) from a variety of environmental templates for use in next-generation 454 pyrosequencing. We hypothesized that, regardless of DNA source, there would be measurable and consistent biases between low- and high-fidelity DNA polymerases as determined by comparing the final sequence-generated community membership. To address this, we selected representative templates derived from sample types studied in microbial ecology research (human feces, lake water, a whole firefly, and cow rumen) and amplified the V5-V8 variable region of the 16S rRNA gene using both low- and high-fidelity DNA polymerases to test the consistency of the final amplicon pools. The V5-V8 region was chosen in order to avoid amplifying homologous regions of the 16S rRNA gene in chloroplast and the 18S rRNA gene in eukaryotes [16]. For two of the templates (human feces and firefly) we performed three separate technical replications.

We sequenced all samples using Roche 454 pyrosequencing, and from these sequences determined: (a) quality and length of sequences, (b) relative detection of community members based on sequence analysis, and (c) similarity among sequence sets by template and technical replicate. In order to examine if bacterial concentration within each template correlated with differences between amplicon pools, we used quantitative PCR (qPCR) to measure the proportion of bacterial DNA within the total DNA used to generate each amplicon library. We found that differences in sequence set composition are not dependent upon polymerase fidelity, but that specific interactions between each polymerase and template influence the obtained

sequence profile. We also provide recommendations for the selection of polymerase type, and for reducing error from technical variability in sequence-based analyses.

Materials and Methods

Sample collection and DNA extraction. All DNA was quantified using a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE) and checked for degradation by gel electrophoresis on a 1 % agarose in TAE gel prior to storage at -20 °C. Sample collection and DNA extraction methods for each sample are detailed as follows.

Rumen: Rumen solids were collected in November of 2011 over the course of three consecutive days at the public Dairy Forage Research Center farm (DFRC, Prairie du Sac, WI), under the authority of the USDA, from a fistulated, lactating Holstein cow as a part of a separate study conducted under the animal use protocol A01104 approved through the University of Wisconsin and the USDA DFRC. All samples were transported on wet ice and frozen at -80 °C prior to DNA extraction. Total genomic DNA was extracted from the solids following a protocol published previously [17] with the following modification: 25:24:1 phenol:chloroform:isoamyl alcohol was used in place of phenol:chloroform. The DNA from each sampling day was pooled in equal concentrations to create a composite DNA sample.

Insect: An adult firefly (*Photinus pyralis*) was collected in June of 2012 at Willow Creek Woods in Madison, WI (43°4′32″N 89°25′30″W). Willow Creek Woods is public property under the authority of the University of Madison-Wisconsin, requiring no permit or approval for sample collection, and *P. pyralis* is not a protected or endangered species. The firefly was immediately placed in ethanol and stored at -20 °C prior to DNA extraction following the method of [18]. Briefly, DNA was extracted using bead disruption and phenol/chloroform. In brief, the insect

was dried at 37 °C to remove the storage ethanol and placed in a 2 mL screw top tube with a single sterilized steel bead (3 mm diameter) and 1 mL of CTAB and mechanically disrupted in a Mini-beadbeater (Biospec Products, Bartlesville, OK) for 2 min, then placed at -80 °C for 2 min 30 sec. This beating-and-freezing was repeated twice, then the debris were pelleted and the supernatant removed and added to 1 ml of 25:24:1 phenol:chloroform:isoamyl alcohol, vortexed, and centrifuged for 5 min max speed in a tabletop centrifuge. The aqueous phase was added to ice-cold isopropanol and gently invert-mixed, then stored at -80 °C for 30 minutes. The total genomic DNA was pelleted by centrifugation for 15 min at 4 °C at max speed in a tabletop centrifuge. After removing the supernatant the pellet was washed with 70 % ethanol and re-suspended in TE.

Fecal: The human fecal sample was collected and processed according to the human use protocol H-2010-0151 as part of a separate research project approved through the University of Wisconsin's Department of Health Sciences Institutional Review Board. Written informed consent from the donor was obtained for use of this sample in research. In brief, the sample was collected within 24 h from a sterile sample cup given to a patient in which they made a deposit and stored at -80 °C until DNA extraction. Fecal DNA was extracted using the QIAamp DNA Stool Mini Kit (Qiagen, Valencia, California) following the manufacturer's protocol with no modifications.

Lake: A freshwater lake sample was collected in September of 2001 from Lake Mendota, Madison, WI (43°4′9″N 89°21′34″W). Lake Mendota is public property, requiring no permit or approval for sample collection. The samples was collected using a sterilized PVC integrated water column sampler that was rinsed twice with lake water from the sample location, and stored in a cooler until it could be filtered. Bacteria were recovered by filtration on 0.2 um

polyethersulfone filters (Pall-Supor-200, Gelman), without pre-filtration. Filters were frozen at -80 ^{o}C and stored prior to DNA extraction as previously described [19]. Briefly, total genomic DNA was extracted from the filter using mechanical disruption in a Mini-BeadBeater (BioSpec Products, Bartlesville, OK) and the FastDNA extraction kit (MP Biomedicals, Solon, OH) using the manufacturer's protocol with minor modifications [20].

Polymerases and reaction conditions. The following polymerases were used for PCR: EconoTaq (Lucigen, Middleton, WI), Platinum Blue PCR Supermix (Invitrogen, San Diego, CA), PfuTurbo (Agilent, Santa Clara, CA), and SlowTaq (Lucigen). Note that SlowTaq is a fusion enzyme between Taq DNA polymerase and a double strand DNA binding domain showing higher affinity than Sso7 type chromatin proteins [21]. The primers 799F-mod6 and 1392R were used to target the V5-V8 variable regions of the 16S rRNA gene while reducing undesirable amplification from insect 18S rRNA and chloroplast 16S rRNA [16], and included the Roche adapters and a 10 bp barcode. The primers were as follows: forward: 5'-AGAGTTTGATCMTGGCTCAG, reverse: 5'- XXXXXXXXXX-GWATTACCGCGGCKGCTG, where XXXXXXXXXX represents the unique 10 bp barcode. All reactions were initially run using manufacturer protocols scaled down to 20 μL reactions with 20 ng genomic DNA and 0.125 μM final concentration of each primer. DNA absent controls were run for each reaction mix using water in place of DNA. Initial cycling conditions were: an initial denaturation of 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 the final extension at 68 °C for 10 min. PCR conditions were modified as needed to achieve amplification as detected by gel electrophoresis of 3 or 10 μL of PCR product on a 1 % agarose TAE gel with ethidium bromide (0.1 mg/mL final concentration in gel). The PCR modifications used were: increased template concentration

(20 - 100 ng per reaction in 20 ng increments); decreased primer concentration (0.0125 μM); addition of 0.4 μL 100 % DMSO (Invitrogen); sample volume (10 – 40 μL in 10 μL increments); each polymerase manufacturer's recommended cycling conditions for annealing and extension parameters; decreasing annealing temperature to 50 °C by gradient; extension times of 1 min and 2 min. Each 20 μL reaction pair was pooled prior to cleaning for removal of primer and small DNA fragment contaminants by gel extraction in 1 % low melting agarose Ultra Pure Aqua*Pōr* LM (National Diagnostics, Atlanta, GA) in TAE, using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irving, CA). DNA was quantified using a Qubit® Fluorometer (Invitrogen, San Diego, CA), then pooled to create a single sample at 1×10^9 molecules per μL based on the average expected amplicon size of 570 bp. An additional round of small fragment removal was performed on this pool using the PureLink Quick PCR Purification Kit (Invitrogen) using the optional buffer system formulated to remove all fragments under 300 bp and following manufacturer guidelines. The column-cleaned DNA pool was re-quantified by Qubit® Fluorometer and diluted to 1×10^6 molecules per μL for use in emPCR and 454 pyrosequencing.

Technical replication of fecal and firefly templates. Single biological replicates of the fecal and firefly DNA templates were subjected to three technical replicates for the following steps: PCR amplification and processing prior to pooling for pyrosequencing. When the product of a single PCR and clean-up protocol was included on multiple pyrosequencing runs, these were not considered full technical replicates and the sequences were pooled for analysis. Except for determination of technical variation, all sequences were pooled by polymerase and template for analysis. All sequencing runs were performed on the same Roche GS Junior.

454 pyrosequencing and data analysis. Amplification and sequencing for all samples was performed following the manufacturer’s guidelines for a Roche 454 GS Junior with the Lib-L kit

and Titanium chemistry. The resulting sequences were screened for quality and analyzed using the bioinformatics software suite mothur [22] v1.33.0, with all samples subjected to identical clean-up and analysis commands. An example generic batch file of the commands used for sequence analysis is presented in Supplemental File S1. Coverage and diversity metrics were calculated without sub-sampling. Inter-group comparisons using rarefaction curves, principal component analysis (PCA) [23], and non-parametric multidimensional scaling (NMDS) [24] were performed in mothur using both entire and sub-sampled sets and included the Morisita-Horn index [25] and Yue and Clayton's theta (Θ_{YC}) [26], both of which measure the degree of community dissimilarity. Intra-group comparisons were performed using sub-sampling to the smallest sequence set in that group. The sequence data sets with identifying barcodes are deposited in the National Center for Biotechnological Information's Short Read Archives under accession number PRJNA245749.

qPCR enumeration of bacterial DNA. A TaqMan probe-base quantitative real-time PCR (qPCR) assay was employed to quantify the abundance of bacterial DNA [27] present in each template using DNA from a pure lab-isolated bacterial culture (*Clostridium perfringens*) as a positive control. Each 25 µl general Bacteria reaction consisted of 12.5 µl Premix Ex Taq (Clontech, Mountain View, CA), 0.4 µM forward primer BAC338F (5'-ACTCCTACGGGAGGCAG-3'), 0.4 µM reverse primer BAC805R (5'-GACTACCAGGGTATCTAATCC-3'), 0.2 µM FAM-labeled probe BAC516F (5'-TGCCAGCAGCCGCGGTAATAC-3'), 8.5 µl sterile H_2O, and 1 ng of DNA template for the experimental wells and 1, 10, 100 or 1000 ng of *C. perfringens* DNA. DNA-absent wells contained the same reaction mixture but with water in place of DNA. Cycling conditions were as follows: initial denaturation at 95 °C for 30 sec, 45 cycles of 95 °C for 5 sec, and 56 °C for 30

File S1. Generic batch file containing all major commands

```
#Supplemental file S1: Generic batch file
#For mothur v1.33.0 64 bit on Ubuntu, C.  McCormick & K.  Jewell 2013-2014
sffinfo(sff=Generic.sff, flow=t, trim=t, fasta=t)
trim.flows(flow=Generic.flow, oligos=Generic.oligos, pdiffs=2, bdiffs=0, fasta=T,
        minflows=450, maxflows=720)
shhh.flows(file=Generic.flow.files)
trim.seqs(fasta=Generic.shhh.fasta, name=Generic.shhh.names, oligos=Generic.oligos, pdiffs=2,
        bdiffs=0, maxhomop=6, minlength=250, flip=T)
count.groups(group=Generic.groups)
unique.seqs(fasta=Generic.fasta, name=Generic.names)
summary.seqs(fasta=Generic.unique.fasta, name=Generic.unique.names)
align.seqs(fasta=Generic.unique.fasta, reference=silva.all.fasta, flip=t)
screen.seqs(fasta=Generic.unique.align, name=Generic.unique.names, group=Generic.groups,
        minlength=300, optimize=end, criteria=90)
count.groups(group=Generic.groups)
filter.seqs(fasta=Generic.unique.good.align, vertical=T, trump=., processors=3)
unique.seqs(fasta=Generic.unique.good.filter.fasta, name=Generic.unique.good.names)
Pre.cluster(fasta=Generic.unique.good.filter.unique.fasta,
        name=Generic.unique.good.filter.names, group=Generic.good.groups, diffs=2)
count.groups(group=Generic.good.groups)
```

```
chimera.uchime(fasta=Generic.unique.good.filter.unique.precluster.fasta,
        name=Generic.unique.good.filter.unique.precluster.names, group=Generic.good.groups,
        processors=3)
remove.seqs(accnos=Generic.unique.good.filter.unique.precluster.uchime.accnos,
        fasta=Generic.unique.good.filter.unique.precluster.fasta,
        name=Generic.unique.good.filter.unique.precluster.names, group=Generic.good.groups)
count.groups(group=Generic.good.pick.groups)
summary.seqs(fasta=Generic.unique.good.filter.unique.precluster.pick.fasta,
        name=Generic.unique.good.filter.unique.precluster.pick.names)
classify.seqs(fasta=Generic.unique.good.filter.unique.precluster.pick.fasta,
        template=nogap.all.fasta, name=Generic.unique.good.filter.unique.precluster.pick.names,
        taxonomy=silva.all.silva.tax, cutoff=80)
remove.lineage(fasta=Generic.unique.good.filter.unique.precluster.pick.fasta,
        name=Generic.unique.good.filter.unique.precluster.pick.names,
        group=Generic.good.pick.groups,
        taxonomy=Generic.unique.good.filter.unique.precluster.pick.silva.wang.taxonomy,
        taxon=Eukaryota)
system(cp Generic.unique.good.filter.unique.precluster.pick.pick.fasta      Generic.final.fasta)
system(cp Generic.unique.good.filter.unique.precluster.pick.pick.names    Generic.final.names)
system(cp Generic.good.pick.pick.groups Generic.final.groups)
system(cp Generic.unique.good.filter.unique.precluster.pick.silva.wang.pick.taxonomy
                Generic.final.taxonomy)
count.groups(group=Generic.final.groups)
```

```
dist.seqs(fasta=Generic.final.fasta, output=lt)
cluster(phylip=Generic.final.phylip.dist, name=Generic.final.names, method=average,
        cutoff=0.15)
make.shared(list=Generic.final.phylip.an.list, group=Generic.final.groups, label=0.05)
summary.single(shared=Generic.final.phylip.an.shared, label=0.05, calc=nseqs-coverage-sobs-
        chao-invsimpson-bergerparker-boneh-efron)
summary.shared(calc=jclass-thetayc-morisitahorn)
classify.otu(list=Generic.final.phylip.an.list, name=Generic.final.names,
        taxonomy=Generic.final.taxonomy, group=Generic.final.groups, label=0.05, cutoff=80,
        basis=otu)
sub.sample(shared=Generic.final.phylip.an.shared, label=0.05, persample=T)
rarefaction.single(shared=Generic.final.phylip.an.0.05.subsample.shared, label=0.05,
        groupmode=F)
tree.shared(shared=Generic.final.phylip.an.0.05.subsample.shared, calc=morisitahorn,
        label=0.05)
pca(shared=Generic.final.phylip.an.0.05.subsample.shared)
dist.shared(shared=Generic.final.subsample.phylip.an.shared, calc=jclass, label=0.05)
nmds(phylip=Generic.final.subsample.final.phylip.an.jclass.0.05.lt.dist)
pca(shared=Generic.final.phylip.an.shared)
nmds(phylip=Generic.final.phylip.dist, iters=1000, mindim=2, maxdim=3)
quit()
```

sec. A plate reading 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 template was run in triplicate on a single plate and compared to a five-member *C. perfringens* DNA standard curve.

Results

Amplification required modification of initial PCR conditions. We selected polymerases of high- or low-fidelity based on a literature survey of pyrosequencing-based studies, and representative publications using these polymerases are given here: PfuTurbo (Agilent) [9,13,28,29], Platinum Blue PCR Supermix (Invitrogen) [30,31,32,33], and EconoTaq (Lucigen) [34,35,36,37]. PfuTurbo is a *Pfu* DNA polymerase mixed with a polymerase-enhancing factor with an error rate six times less than *Taq* (Agilent product insert); Platinum Blue PCR Supermix is a recombinant *Taq* DNA polymerase with an error rate equal to *Taq* (Invitrogen product insert); EconoTaq is a *Taq* DNA polymerase (Lucigen product insert). Finally, we chose to include SlowTaq (Lucigen), a non-commercial polymerase that, to our knowledge, has not previously been used for pyrosequencing applications. Because of its reported high fidelity, coupled with its tenacity of template DNA binding during PCR [21], it is expected to lower the rate of PCR-generated chimera formation and error.

Final PCR modifications are given in Table 1. Amplification was considered successful when a product band was visible under UV light when 10 μL of PCR product were run on a 1% agarose TAE gel with ethidium bromide, independent of the formation of a band corresponding to primer-dimer formation. If no product band was visible alterations were made to the PCR conditions as detailed in the methods section. A highly successful amplification was defined as

Table 1. Specific PCR parameters for 20 ng of template in 20 µL reactions.

Template	Polymerase	DMSO	Primer (µM each)
Fecal	EconoTaq	-	0.0125
Firefly	EconoTaq	NA	NA
Lake	EconoTaq	-	0.0125
Rumen	EconoTaq	-	0.0125
Fecal	Platinum Blue	-	0.0125
Firefly	Platinum Blue	+	0.1250
Lake	Platinum Blue	-	0.1250
Rumen	Platinum Blue	-	0.1250
Fecal	PfuTurbo	-	0.1250
Firefly	PfuTurbo	-	0.1250
Lake	PfuTurbo	-	0.1250
Rumen	PfuTurbo	-	0.1250
Fecal	SlowTaq	-	0.1250
Firefly	SlowTaq	-	0.1250
Lake	SlowTaq	-	0.1250
Rumen	SlowTaq	-	0.1250

NA - No amplification detected by gel electrophoresis using any combination of parameters as described in the Methods; (+) - addition of 0.4 µL of 100 % DMSO; (-) - No DMSO

one where the sole or major product was the desired amplicon. The factor that most improved amplification was a decrease in total primer concentration, which resulted in amplicon formation over primer-dimer formation as visualized by agarose gel. Changing cycling conditions and increasing template concentration did not improve amplification success. No combination of variable parameters resulted in amplification of the firefly DNA by EconoTaq, as determined by visualization on an agarose gel.

Sequencing and coverage results. All successful amplifications were subjected to 454 pyrosequencing, generating a total of 173,822 sequences, of which 161,174 passed all filter and quality control steps with a mean length of 449 bp. Sequence counts and diversity metrics (calculated without sub-sampling) are presented in Table 2, showing a high degree of estimated community coverage (Good's [38], where 1.00 is total coverage) of 0.87 - 1.00 and a leveling appearance of all rarefaction curves, even after sub-sampling (Fig. S1). Variation within each template by polymerase, in terms of the number of total or rare detected genus-level operation taxonomic units (OTUs, 95% sequence similarity [39]) (Table 2), fluctuated by less than 15 % of the mean for all but Firefly (which was nearly 35 % in both cases). For all analyses, we classified each sequence to an OTU, and also binned the OTUs by phylum. We found that phyla detection was not consistent within each template (Table 3).

Chimera count, but not amplicon length, was impacted by polymerase selection. Given that polymerase fidelity and quality could impact the production of template suitable for usage in 454 pyrosequencing, it was hypothesized that DNA polymerases with higher fidelity and tenacity would reduce the number of sequences discarded due to low quality (length and base-call accuracy) and/or chimeras. To determine if this correlation existed, all samples were parsed for the percentage discarded due to short length and chimeric status. The results of this analysis are

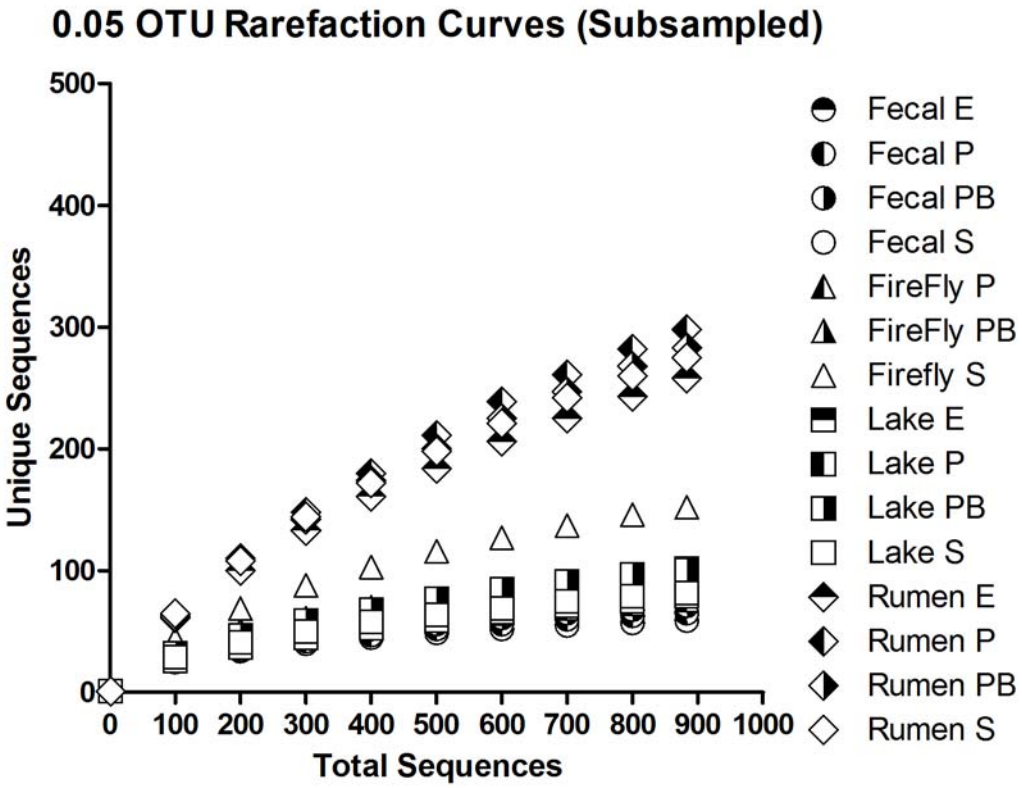

Figure S1. Sub-sampled rarefaction curves using genus-level OTUs (95 % sequence similarity) for all samples pooled by polymerase and template. (E) EconoTaq, (P) PfuTurbo, (B) Platinum Blue, (S) SlowTaq.

Table 2. Sequencing distribution metrics for all templates before sub-sampling.

Template	Polymerase	Sequences	S_{obs} [a]	Rare S_{obs} [b]	Good's coverage	Chao1	Inverse Simpson
Fecal	EconoTaq	50,046	261	246	1.00	475.63	6.99
Fecal	PfuTurbo	10,925	266	250	0.99	619.91	8.29
Fecal	Platinum Blue	13,600	200	182	1.00	267.00	9.86
Fecal	SlowTaq	17,136	232	215	0.99	422.38	9.96
Firefly	PfuTurbo	14,518	178	168	1.00	237.91	6.31
Firefly	Platinum Blue	20,525	280	265	1.00	324.68	9.64
Firefly	SlowTaq	23,605	366	350	1.00	409.75	10.30
Lake	EconoTaq	1,609	108	93	0.97	189.67	7.39
Lake	PfuTurbo	912	103	89	0.94	230.15	8.16
Lake	Platinum Blue	1,206	115	100	0.96	162.00	10.16
Lake	SlowTaq	1,174	100	83	0.96	157.50	7.54
Rumen	EconoTaq	1,935	390	375	0.90	642.89	34.32
Rumen	PfuTurbo	883	298	287	0.81	565.38	37.88
Rumen	Platinum Blue	1,264	351	336	0.86	579.12	29.09
Rumen	SlowTaq	1,836	414	400	0.89	634.25	45.97

[a] S_{obs} and genus-level OTU counts are equal

[b] Defined as OTUs with < 1.00% relative abundance

Table 3. Distribution of all phyla by percent relative sequence abundance.

	Fecal				Firefly			Lake				Rumen			
Phylum	**B**	**E**[a]	**P**	**S**	**B**	**P**	**S**	**B**	**E**	**P**	**S**	**B**	**E**	**P**	**S**
Acidobacteria		[b]			-[c]	0.05	-								
Actinobacteria	0.51	0.4	0.22	0.41	23.7	15.1	11.0	53.2	59.4	60.1	58.2	0.40	0.05	0.68	-
Armatimonadetes					1.09	-	-	0.27	0.07	0.47	0.27				
Bacteroidetes	12.0	14.4	14.0	14.8	6.67	2.97	14.2	12.0	18.0	8.95	20.1	39.4	43.7	33.3	47.5
BRC1					0.01	-	-								
Chlorobi	0.07	0	0.01	-	0.49	-	0.22	-	-	0.12	-				
Chloroflexi	0.01	-	-	-	0.52	0.79	0.56	5.14	3.53	5.58	3.30	-	0.1	0.11	-
Elusimicrobia					-	0.22	-					-	-	-	0.11
Fibrobacteres	0.029	0.022	0.009	0.02	0.94	30.8	29.2					0.79	0.88	1.81	0.6
Firmicutes	78.9	77.8	75.3	76.5	10.9	32.6	30.1					48.0	44.0	46.4	38.6
Fusobacteria	0.03	0.01	0.02	0.02	0.04	-	0.01								
Gemmatimonadetes								-	-	-	0.09				
GN02								-	-	-	0.09	-	0.05	-	-
Lentisphaerae												-	0.05	0.23	-
Planctomycetes	0.02	-	0.02	-	0.01	0.57	0.10	0.54	0.78	0.7	0.18	0.24	-	0.11	0.05
Proteobacteria	3.48	3.49	4.24	4.24	55.2	15.5	10.6	24.6	15.0	19.2	15.3	6.97	7.86	11.8	7.36
Spirochaetes	0.01	0	-	-	0.13	0.73	0.3					1.74	1.09	3.06	1.58
SR1					-	-	0.01					0.40	0.26	0.11	1.09
Synergistetes												0.08	-	0	0.11
Tenericutes	1.04	0.73	0.78	1.01	0.08	0.01	0.66					1.19	1.4	1.13	2.07
Thermi					-	-	0.72								
TM6								-	-	0.12	-				
TM7					0.15	0.03	0.16	-	-	0.12	-	0.48	0.21	0.45	0.55
Verrucomicrobia	3.97	3.14	5.32	2.95	0.11	0.33	1.66	3.78	3.20	3.6	2.23	-	0.26	0.23	-
Unclassified	-	0.01	0.13	0.03	0.02	0.3	0.45	0.54	-	1.05	0.27	0.40	0.05	0.57	0.38
Total phyla	12	11	11	9	16	14	16	8	7	11	10	12	14	15	12

[a] E = EconoTaq, P = PfuTurbo, B = Platinum Blue, S = SlowTaq

[b] Blank cells indicate no detection of that phylum for that template type.

[c] Dashes (-) indicate no detection of that phylum for that combination of template and polymerase.

summarized in Table 4. The removal of short and chimeric sequences did not depend on polymerase within each template ($P > 0.05$), and both metrics were highly variable for each polymerase between templates. In all but the Rumen sample, the low-fidelity polymerases (Platinum Blue and EconoTaq) had the lowest number of removed sequences in both conditions (Table 4).

Fecal, Lake, and Rumen sequence sets clustered tightly, but sets of identical OTUs are small. The total sequence sets, treated as 95 % OTUs, for the Fecal, Lake, and Rumen templates formed tight clusters when analyzed as entire sets or as sub-samples using principal component analysis (PCA) and non-parametric multidimensional scaling (NMDS) (Fig. 1); the Firefly sequence sets did not cluster, and are discussed separately below. We analyzed each set of sequences by template origin using a standard bacterial taxonomic sequence database (Green Genes [40]), and by non-taxonomic comparisons of sequence diversity (diversity, richness, and total-community analyses). Within each sample, the percentage of genus-level OTUs shared among all polymerases (Fecal 22.7 %, Firefly 9.4 %, Lake 17.5 %, Rumen 11.4 %) was less than those unique to one or more polymerases (Fig. 2).

Detection of the firefly bacterial community between replicates and polymerases was highly variable. The Firefly template was the only sample from which we were unable to amplify using all polymerases (Table 1). Moreover, we did not see clustering of total bacterial sequence sets (Fig. 1), and found an extremely high degree of variation among OTUs binned to phyla (Table 3) in terms of both presence and abundance. Even when the same OTU was detected in all samples, the relative abundance was not consistent. For example, a single OTU varied in sequence abundance from 0.4 % for SlowTaq to 0.7 % for PfuTurbo, and to 27.1 % for Platinum

Table 4. Comparison of sequence quality for each template by polymerase.

Template	Polymerase	Removed short (%)[a]	Removed chimeras (%)
Fecal	EconoTaq	1.55	0.62
Fecal	PfuTurbo	5.08	22.06
Fecal	Platinum Blue	1.12	0.59
Fecal	SlowTaq	2.23	6.83
Firefly	PfuTurbo	3.95	0.75
Firefly	Platinum Blue	2.26	0.00006
Firefly	SlowTaq	3.95	0.24
Lake	EconoTaq	10.85	2.85
Lake	PfuTurbo	12.25	18.46
Lake	Platinum Blue	7.66	14.24
Lake	SlowTaq	10.99	6.50
Rumen	EconoTaq	4.14	5.71
Rumen	PfuTurbo	7.28	9.47
Rumen	Platinum Blue	6.75	9.65
Rumen	SlowTaq	4.03	11.50

[a] Sequences were considered to be too short if they were less than 250 bp (*trim.seqs* command).

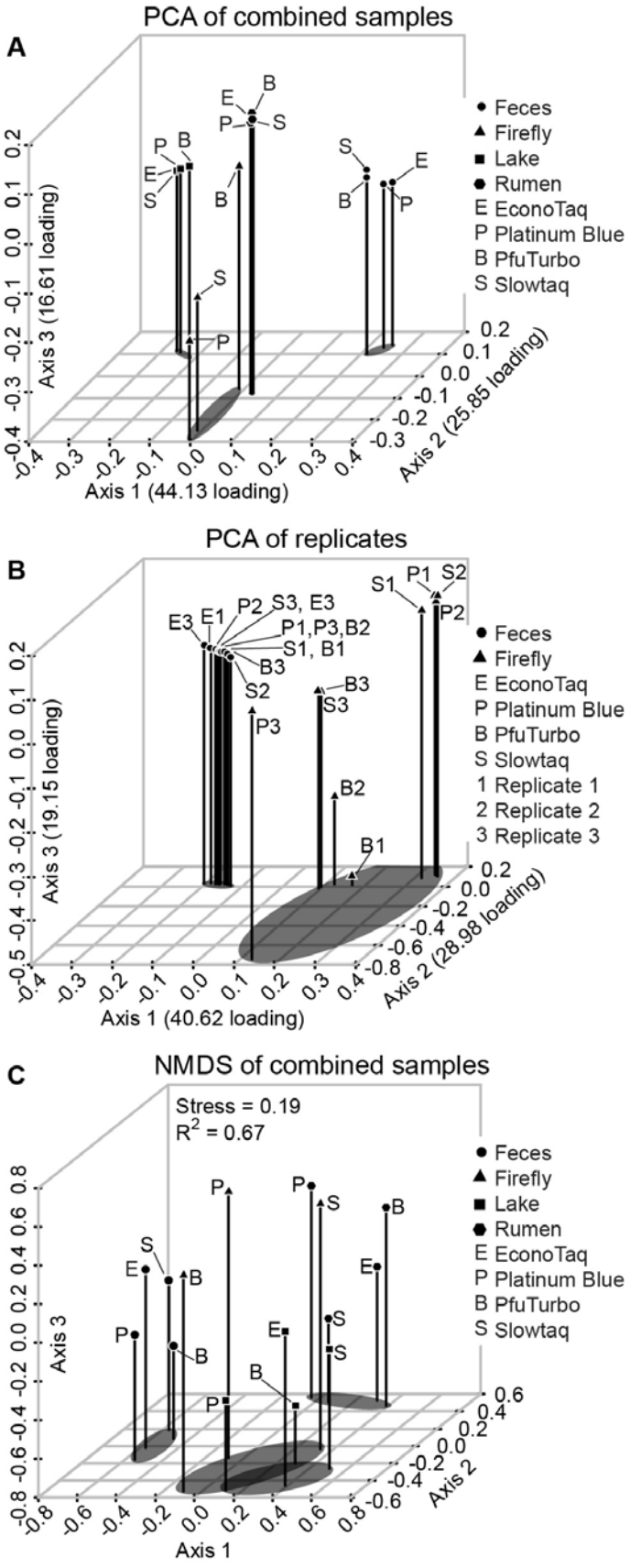

Figure 1. Total sequence set structure comparisons for both pooled and replicated samples using sub-sampling. **(A)** Three-dimensional PCA of combined samples. **(B)** Three-dimensional PCA of technical replicates for Fecal and Firefly. Individual technical replicates are identified by number. **(C)** Three-dimensional NMDS of combined samples. Analysis was repeated with non-sub-sampled sets with no significant differences in clustering patterns or locations. Polymerase abbreviations: (E) EconoTaq, (P) PfuTurbo, (B) Platinum Blue, (S) SlowTaq. Templates are identified by symbol and color: (brown circle) Fecal, (yellow triangle) Firefly, (red hexagon) Rumen, (blue square) Lake.

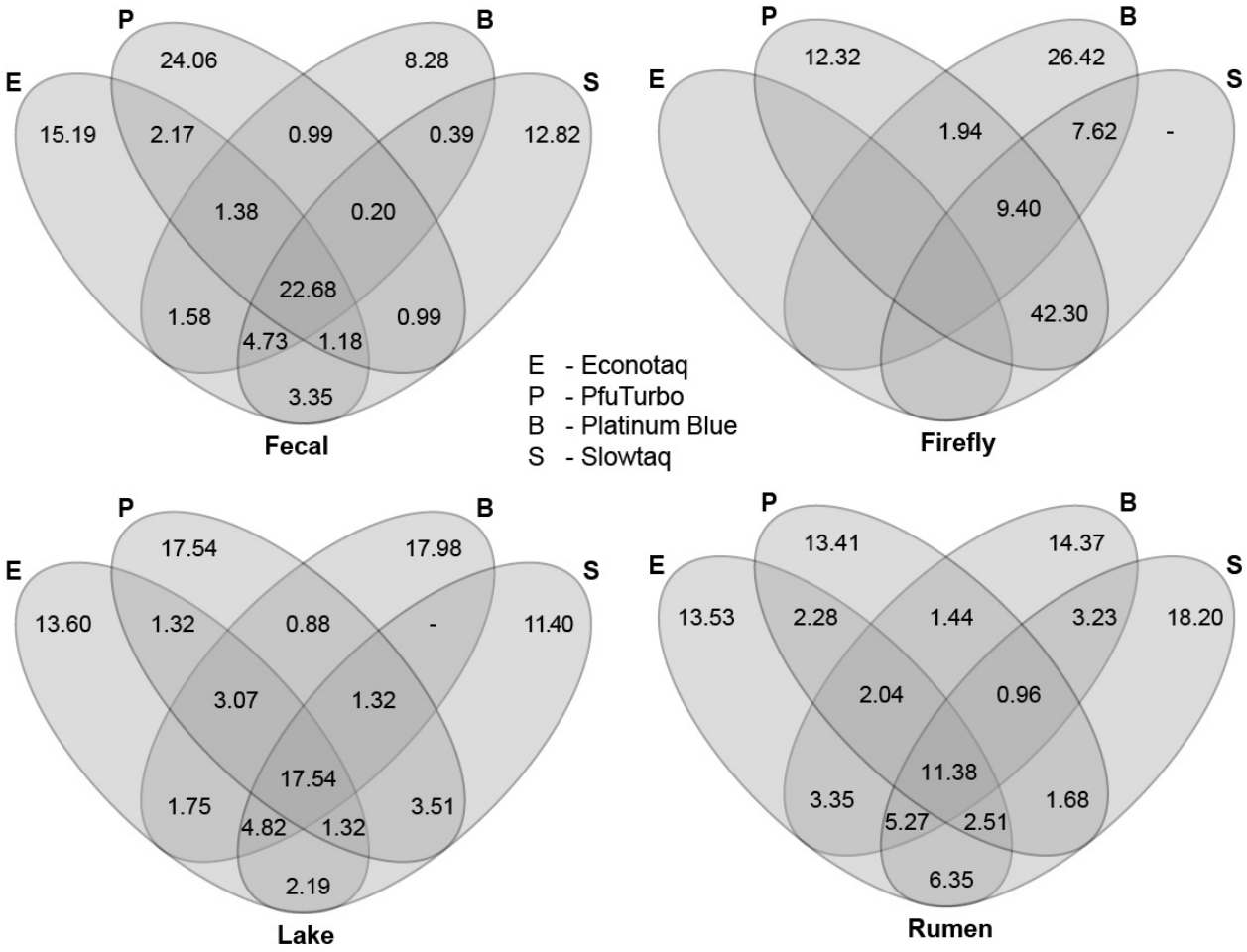

Figure 2. Venn diagrams of percent relative abundance of genus-level OTUs unique to each intersection identified by polymerase abbreviation and community origin. Numbers in bold highlight the shared percentage of identical OTUs. (E) EconoTaq, (P) PfuTurbo, (B) Platinum Blue, (S) SlowTaq, (-) no OTUs unique to that condition.

Blue. This compounded the differences between the detected sequence sets due to the presence/absence of OTUs.

We then tested whether technical variation in amplifying or sequencing of the Firefly sample was responsible for the high degree of observed dissimilarity. We performed three full technical replicates of the Firefly (a highly dissimilar sequence set) and Fecal (a highly similar sequence set) templates. We found that the degree of variation, as measured by PCA clustering (Fig. 1B) and similarity metrics (Table 5), was minimal for the Fecal template, but high for the Firefly template. Within the technical replicates, the shared, identical OTUs (of any abundance) were improved over that seen between polymerases for Fecal (19.3 % for PfuTurbo, 27.7 % for SlowTaq, 36.8 % for Platinum Blue, and 40.0 % for EconoTaq) but not for the Firefly (4.0 % for Platinum Blue, 6.2 % for PfuTurbo, and 6.9 % for SlowTaq).

A comparison of the replicates revealed that, although the richness was similar (as determined by the Jaccard similarity coefficient), the final sequence set and structure was dissimilar (determined by Morisita-Horn index and Yue and Clayton theta coefficient, Θ_{YC}) (Table 5). The Morisita-Horn index is independent of sample size but systematically underestimates similarity [41], while Θ_{YC} uses weighted proportions of unshared and shared OTUs in addition to sequence abundance [26]; both measure the degree of community dissimilarity, and as implemented in mothur, returns a value between 0 and 1, with 1 indicating complete dissimilarity. The variation in the Firefly polymerase replicates, and for the pooled Firefly datasets, was only surpassed by the dissimilarity observed between the templates (0.90 - 1.00 by Morisita-Horn and 0.95 - 1.00 by Θ_{YC}).

Using minimum sequence- or OTU-based cut-offs reduced polymerase-based variation. The observed dissimilarity within each template was driven by large numbers of low-abundance

Table 5. Comparison of detected community similarity for genus-level OTUs.

Template	Polymerase	Jaccard[a]	SDM[b]	Morisita-Horn[c]	SDM	Θ_{YC}[d]	SDM
Fecal	EconoTaq	0.47	0.04	0.04	0.03	0.07	0.05
	PfuTurbo	0.71	0.02	0.01	0.01	0.03	0.02
	Platinum Blue	0.51	0.04	0.02	0.01	0.04	0.01
	SlowTaq	0.59	0.08	0.01	0.01	0.03	0.01
	All	0.59	0.01	0.03	0.02	0.05	0.04
Firefly	PfuTurbo	0.83	0.06	0.67	0.06	0.69	0.53
	Platinum Blue	0.86	0.08	0.71	0.08	0.78	0.34
	SlowTaq	0.83	0.02	0.58	0.02	0.63	0.51
	All	0.81	0.01	0.61	0.45	0.74	0.39
Lake	All	0.63	0.04	0.03	0.02	0.05	0.03
Rumen	All	0.71	0.05	0.08	0.03	0.15	0.04

[a] Mean Jaccard similarity coefficient using observed sequence set richness

[b] SDM = Standard deviation of the mean

[c] Mean Morisita-Horn index using sequence set community structure

[d] Mean Yue and Clayton theta (Θ_{YC}) dissimilarity measurement using sequence set community structure

OTUs (<1.00 % relative sequence abundance); when comparing only OTUs with ≥1.00 % relative sequence abundance in at least one condition, the shared percentages increased dramatically to 100 % for Fecal (18 OTUs), 75.0 % for Firefly (21 of 28 OTUs), 94.7 % for Lake (18 of 19 OTUs), and 87.0 % for Rumen (20 of 23 OTUs). Because many OTUs may represent important community members that are under the arbitrary 1.00 % cut-off, we also employed a less-stringent requirement that each OTU have two or more sequences in each combination of template and polymerase. Using this second set of parameters with the replicated Fecal dataset, we retained 95.5 % of the Fecal sequences in 39 OTUs (of 529) as the shared set across all enzymes and technical replicates. Similar improvements were observed for the Rumen and Lake samples: Rumen with 80.3 % sequences retained in 54 OTUs (of 845), and Lake with 94.0 % sequences retained in 27 OTUs (of 228). The Firefly, with its extreme variation in sequencing results, retained only 10.3 % of sequences in 4 OTUs (of 617).

The concentration of bacterial DNA was not consistent across templates. We used qPCR to measure the relative abundance of 16S rRNA gene copies present in each template. Our positive control was DNA extracted from a lab-isolated pure bacterial culture (*Clostridium perfringens*), from which we generated a ten-fold dilution standard curve in triplicate. We calculated the proportion of bacterial DNA in each experimental template using the best-fit line for our five-point control dilution ($R^2 = 0.999$). Our calculations showed that there was a high degree of variation by template in the proportion of bacterial DNA present. Based on our qPCR data, of the 20 ng of total DNA used as input for amplification to create amplicon pools for sequencing, the Rumen sample contained 9.4 ± 0.5 ng, the Lake sample 2.1 ± 0.1 ng, and the Feces sample 1.3 ± 0.1 ng of bacterial 16S rRNA genes. The Firefly sample had a concentration barely distinguishable from our qPCR DNA absent negative controls (0.0005 ± 0.00003 ng for the

Firefly and 0.0004 ± 0.0001 for DNA absent). It should be noted that visible bands were observed from our Firefly amplicon library-generating PCR, whereas no visible bands were seen from our DNA absent control reactions, leading us to assume that the sequences generated from the Firefly were not due to delayed amplification of lab-borne DNA contamination.

Discussion

The application of large scale, high-throughput sequencing technology has become increasingly common in biological research over the past decade. However, there still exists much disagreement as how to best prepare or interpret the massive data sets generated by these methods. For microbial ecology, it is known that the methods used in collecting, preparing, and analyzing samples can greatly influence the ability to accurately sequence the original community [6,7,14,39]. Because most library preparation methods rely upon an initial PCR amplification of template DNA, GC-bias and polymerases fidelity can materially impact the coverage and quality of DNA pools used in sequencing reactions [9]. This is underscored by the finding that polymerase selection can be used to target specific sets of DNA [42]. Unintentional polymerase-introduced bias is particularly worrisome for any research drawing conclusions from large-scale comparisons of datasets generated using disparate methods, since the methods themselves have the possibility of being significant factors in the dissimilarity of community detection. We chose to examine the degree of impact resulting from polymerase selection used during library preparation in order to quantify the effects of polymerase fidelity (low or high) and the rate of processivity on the consistency of sequencing results. To determine if there were consistent patterns due to polymerase selection across multiple types of original DNA pools, we chose to use four environmentally-derived DNA templates highly represented in the current

microbial ecology literature: water [13,43,44,45], human feces [1,8,46], ruminal contents [47,48,49,50], and insects [16,51,52].

We found that the ease of amplification varies by template and polymerase, with the low-fidelity polymerases requiring multiple modifications from our initial PCR protocol. Despite requiring modifications, the EconoTaq and the Fecal/Platinum Blue samples did not stand apart when assessed by total community clustering, similarity or dissimilarity measures, or number of unique OTUs detected (except for the Firefly, for which variation was high between all sequence sets). Unlike Hurwitz *et al.* [13], who reported that PfuTurbo detected fewer rare sequences (< 1 % of total sequences) as compared to another high-fidelity polymerase (TaKaRa), we found that PfuTurbo detected more unique genus-level OTUs for the Feces, Lake, and Rumen templates (Fig. 2). We suggest that for projects using multiple templates, small-scale preliminary amplification be performed with multiple high-fidelity polymerases to determine which polymerase can easily create a high-concentration PCR product under a single set of conditions.

A key attribute of 16S rRNA sequence sets is the formation of chimeric sequences. Chimeras are formed when DNA synthesis is interrupted and the partial amplicon anneals to new template sharing localized homology (such as the conserved regions of the 16S rRNA gene) to the original sequence [53], and can result in significant sequence loss during filtering [37,54,55,56]. In our study the number of detected chimeras varied for each polymerase by template, although the lowest percentage of detected chimeras (Table 4) was always from a low-fidelity polymerase with normal processivity. It is possible that factors such as the complexity of template mixture, hold-over compounds from the DNA extraction method, or DNA secondary structure, may strongly influence chimera formation during initial PCR. We suggest that, for projects using a template of a type shown to have a potentially high rate of chimera formation (as

we observed for DNA derived from a freshwater system), and that do not require a high-fidelity polymerase for routine amplification, a low-fidelity polymerase be selected in order to avoid losing high numbers of sequences to chimera-filtering algorithms.

We also found that a low number of all OTUs were identical between polymerases (the highest being Fecal at 22.7 %) and technical replicates (19.3 - 40.0 % for Fecal). This was driven in large part by OTUs composed of singlet sequences. Even when all OTUs were binned at the phylum level, which should reduce the impact of single genus-level, low-abundance OTUs, we found that any phylum with 0.26 % relative sequence abundance (Rumen sample, phylum Verrumicrobia) or less was often not detected by one or more polymerases for that same template. We acknowledge that using any percent cut-off limits the detection of potentially important low-abundance OTUs, reduces the apparent complexity of each system, and potentially introduces Type II errors (increased false negatives). However, by focusing on OTUs with more than one sequence in every sample, we dramatically improved the shared set of OTUs to 80.3 - 95.5 % (excluding Firefly). Based on these observations, we suggest that for studies combining datasets generated using different methods (such as meta-analyses of multiple published datasets), phyla with less than 0.3 % total sequence abundance, and OTUs not represented by more than one sequence in any replicated samples, be treated cautiously when drawing meaningful conclusions; these sequences should ideally be removed entirely from the analyses. We further propose that multiple sequence and OTU abundance cut-off values be used in order to test the robustness of any conclusions drawn from community-level differences.

Technical variation due to repeated PCR amplification [15] and between 454 pyrosequencing runs [50] from a single sample has been shown to be negligible. Based upon a previous, large-scale study examining the impact of technical and location-originating

sequencing error using the GS FLX 454 pyrosequencing platform [12], we expected to observe a Yue and Clayton theta dissimilarity (Θ_{YC}) near to 0.049 between replicates. Indeed, we found a relatively small degree of dissimilarity in our replicated Fecal sample (Θ_{YC} = 0.03 - 0.07) and between replicates from different polymerase treatments.

This was in contrast to our Firefly sample (Θ_{YC} = 0.63 - 0.78), and we conclude that the observed high dissimilarity was not due to technical error or polymerase choice, as there was no clear clustering by either pooled or technically replicated samples (Fig. 1). One possibility was the complicating presence of diet-derived chloroplast sequences and host DNA retained during whole-host extraction of DNA from insects [16], but we did not retain chloroplast or host sequences in our final data sets. Instead, we posit that variability in the Firefly sample was due to the extremely low abundance of bacterial DNA as detected by qPCR. Extremely low abundance of target template (below a single ng in each reaction) may have resulted in dramatically increased bias in the first few rounds of PCR amplification, further exacerbated by the possible complete absence of some bacterial representatives in any given template aliquot. Therefore, any template of potentially low bacterial DNA concentration (such as whole-host homogenates) may have erratic detection of the bacterial community. In such cases, we recommend that the amount of bacterial DNA present in the sample be determined by qPCR in order to ensure a high abundance of bacterial DNA in PCR amplification.

In general, our broad suggestion is that polymerase selection be informed by (a) ease of amplification, (b) probable chimera formation, and (c) future comparison of datasets. For studies where one or more templates are difficult to amplify, we recommend that a high-fidelity polymerase by used so that template-specific PCR modifications are minimized. For studies using a template type previously shown to result in a high number of sequences lost to chimera-

detection software, we recommend that a low-fidelity polymerase be used to minimize such losses.

For groups working on multiple projects that wish to combine or compare their datasets, we recommend that a single polymerase be selected and used among them. When comparing data sets generated using different library preparation methods, measures, such as stringent relative abundance cut-offs or removal of OTUs with single representative sequences, be taken to minimize false discovery rates. Although not yet feasible for most 16S rRNA based research, we reiterate the recommendation of others [9,10] that ideally no amplification steps be performed in preparing samples for sequencing, as any actions taken upon a DNA pool will have some inherent bias. Finally, we express the hope that as large-scale sequence data sets become increasingly common, there will be further work done to generate a consensus of methods and analyses appropriate to their usage and interpretation.

Acknowledgements

We thank D. Mead (Lucigen) for providing SlowTaq polymerase. We thank the Wisconsin Youth Apprenticeship Program in Biotechnology and Oregon High School for providing training for E. Strassman and C. McCormick. We thank A. Andersen for performing DNA extractions from human stool samples, E. Houck for assistance in performing pyrosequencing runs, and T. Neumann for assistance with qPCR. We thank all members of the Suen lab for their support, insightful discussions, and careful reading of this manuscript. This work was funded by the University of Wisconsin-Madison College of Agriculture and Life Sciences. KJ was funded by a USDA NIFA Fellowship Grant #2012-0119. DNA from human stool samples were obtained through a study funded by the Wisconsin Center for Infectious Diseases, Medical Education and

Research Committee, UW School of Medicine and Public Health. Additional support was provided by the United States National Science Foundation (KDM, MCB-0702395 and NTL-LTER DEB-0822700: CRC MCB-0702025). This study was also supported by the Lucigen Corporation, which donated the SlowTaq DNA polymerase. The funders listed above had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Lucigen Corporation provided support for this study through the donation of the pre-production SlowTaq DNA polymerase. There are no products in development or marketed products to declare associated with this study.

References

1. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, et al. (2009) Bacterial community variation in human body habitats across space and time. Science 326: 1694-1697.
2. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, et al. (2008) Functional metagenomic profiling of nine biomes. Nature 452: 629-632.
3. Loman N, Constantinidou C, Chan J, Halachev M, Sergeant M, et al. (2012) High-throughput bacterial genome sequencing: an embarassment of choice, a world of opportunity. Nature Reviews Microbiology 10: 599-606.
4. Nordfors L, Jansson M, Sandberg G, Lavebratt C, Sengul S, et al. (2002) Large-scale genotyping of single nucleotide polymorphisms by Pyrosequencing™ and validation against the 5′nuclease (Taqman®) assay. Human Mutation 19: 395-401.
5. Whitford MF, Forster RJ, Beard CE, Gong J, Teather RM (1998) Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA Genes$^{\beta}$. Anaerobe 4: 153-163.
6. Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, et al. (2013) Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. PLOS ONE 8: e74787.
7. Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, et al. (2011) Accessing the Soil Metagenome for Studies of Microbial Diversity. Applied and Environmental Microbiology 77: 1315-1324.
8. Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, et al. (2014) The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing. PLOS ONE 9: e88982.
9. Quail MA, Otto TD, Gu Y, Harris SR, Skelly TF, et al. (2012) Optimal enzymes for amplifying sequencing libraries. Nat Meth 9: 10-11.
10. van Dijk EL, Jaszczyszyn Y, Thermes C (2014) Library preparation methods for next-generation sequencing: Tone down the bias. Experimental Cell Research 322: 12-20.
11. Schloss PD, Gevers D, Westcott SL (2011) Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. PLOS ONE 6: e27310.
12. Schloss PD, Schubert AM, Zackular JP, Iverson KD, Young VB, et al. (2012) Stabilization of the gut microbiome following weaning. Gut Microbes 3: 383-393.
13. Hurwitz BL, Deng L, Poulos BT, Sullivan MB (2013) Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. Environmental Microbiology 15: 1428-1440.
14. Wintzingerode FV, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiology Reviews 21: 213-229.
15. Huse SM, Young VB, Morrison HG, Antonopoulos DA, Kwon J, et al. (2014) Comparison of brush and biopsy sampling methods of the ileal pouch for assessment of mucosa-associated microbiota of human subjects. Microbiome 2.
16. Hanshew AS, Mason CJ, Raffa KF, Currie CR (2013) Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. Journal of Microbiological Methods 95: 149-155.

17. Stevenson DM, Weimer PJ (2007) Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Applied Microbiology and Biotechnology 75: 165-174.
18. Hillis D, Marble B, Larson A, Davis S, Zimmer E (1996) Nucleic Acids IV: Sequencing and Cloning; Hillis D, Moritz C, Marble B, editors. Sunderland, MA: Sinauer Associates, Inc.
19. Shade A, Kent AD, Jone SE, Newton RJ, Triplett EW, et al. (2007) Interannual dynamics and phenology of bacterial communities in a eutrophic lake. Limnology and Oceanography 52: 487-494.
20. Yannarell AC, Kent AD, Lauster G, Kratz T, Triplett EW (2003) Temporal patterns in bacterial communities in three temperate lakes of different trophic status. Microbial Ecology 46: 391-405.
21. Mead D (2014): Lucigen.
22. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing MOTHUR: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75: 7537-7541.
23. Gower J (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 325-338.
24. Borg I, Groenen P (1997) Modern Multidimensional Scaling. New York: Springer.
25. Horn H (1966) Measurement of "overlap" in comparative ecological studies. The American Naturalist 100: 419-424.
26. Yue J, Clayton M (2005) A similarity measure based on species proportions. Communications in Statistics - Theory and Methods 34: 2123-2131.
27. 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. Biotechnology and Bioengineering 89: 670-679.
28. Lee OO, Wang Y, Yang J, Lafi FF, Al-Suwailem A, et al. (2011) Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. ISME Journal 5: 650-664.
29. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, et al. (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. Proceedings of the National Academy of Sciences 103: 12115-12120.
30. de Oliveira MNV, Jewell KA, Freitas FS, Benjamin LA, Tótola MR, et al. (2013) Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. Veterinary Microbiology 164: 307-314.
31. Fierer N, Hamady M, Lauber CL, Knight R (2008) The influence of sex, handedness, and washing on the diversity of hand surface bacteria. Proceedings of the National Academy of Sciences 105: 17994-17999.
32. Tiao G, Lee CK, McDonald IR, Cowan DA, Cary SC (2012) Rapid microbial response to the presence of an ancient relic in the Antarctic Dry Valleys. Nature Communications 3: 660.
33. Wong CNA, Ng P, Douglas AE (2011) Low-diversity bacterial community in the gut of the fruitfly Drosophila melanogaster. Environmental Microbiology 13: 1889-1900.
34. Berry D, Ben Mahfoudh K, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. Applied and Environmental Microbiology 77: 7846-7849.

35. Humblot C, Guyot JP (2009) Pyrosequencing of tagged 16S rRNA gene amplicons for rapid deciphering of the microbiomes of fermented foods such as pearl millet slurries. Applied and Environmental microbiology 75: 4354-4361.
36. Keller I, Taverna A, Seehausen O (2010) A pyrosequencing assay for the rapid discrimination of mitochondrial lineages in the Salmo trutta species complex. Molecular Ecology Resources 11: 196-199.
37. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. Nat Meth 6: 639-641.
38. Good I (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40: 237-264.
39. Schloss PD (2010) The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies. PLoS Comput Biol 6: e1000844.
40. DeSantis T, Hugenholtz P, Larsen N, Rojas M, Brodie E, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench campatible with ARB. Applied and Environmental Microbiology 72: 5069-5072.
41. Chao A, Chazdon RL, Colwell RK, Shen T-J (2006) Abundance-based similarity indices and their estimation when there are unseen species in samples. Biometrics 62: 361-371.
42. Dabney J, Meyer M (2012) Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. BioTechniques 52: 87-94.
43. Herlemann DPR, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, et al. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME Journal 5: 1571-1579.
44. Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, et al. (2010) Viral and microbial community dynamics in four aquatic environments. ISME Journal 4: 739-751.
45. Youngblut N, Shade A, Read J, McMahon KD, Whitaker J (2013) Lineage-specific response to environmental change in microbial communities. Applied and Environmental Microbiology 79: 39-47.
46. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, et al. (2008) Evolution of mammals and their gut microbes. Science 320: 1647-1651.
47. Pitta DW, Kumar S, Veiccharelli B, Parmar N, Reddy B, et al. (2014) Bacterial diversity associated with feeding dry forage at different dietary concentrations in the rumen contents of Mehshana buffalo (Bubalus bubalis) using 16S pyrotags. Anaerobe 25: 31-41.
48. Gruninger RJ, Sensen CW, McAllister TA, Forster RJ (2014) Diversity of Rumen Bacteria in Canadian Cervids. PLOS ONE 9: e89682.
49. Mullins CR, Mamedova LK, Carpenter AJ, Ying Y, Allen MS, et al. (2013) Analysis of rumen microbial populations in lactating dairy cattle fed diets varying in carbohydrate profiles and Saccharomyces cerevisiae fermentation product. Journal of Dairy Science 96: 5872-5881.
50. Kittelmann S, Seedorf H, Walters W, Clemente JC, Knight R, et al. (2013) Simultaneous amplicon sequencing to explore co-occurence patterns of bacterial, archaeal, and eukaryotic microorganisms in rumen microbial communities. PLOS ONE 8: e47879.
51. Kautz S, Rubin BER, Russell JA, Moreau CS (2013) Surveying the microbiome of ants: comparing 454 pyrosequencing with traditional methods to uncover bacterial diversity. Applied and Environmental Microbiology 79: 525-534.

52. Jing X, Wong ACN, Chaston JM, Colvin J, McKenzie CL, et al. (2014) The bacterial communities in plant phloem-sap-feeding insects. Molecular Ecology 23: 1433-1444.
53. Kopczynski ED, Bateson MM, Ward DM (1994) Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. Applied and Environmental Microbiology 60: 746-748.
54. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environmental Microbiology 12: 118-123.
55. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2006) New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. Applied and Environmental Microbiology 72: 5734-5741.
56. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Research 21: 494-504.

APPENDIX THREE

List of publications not included in this thesis produced from work performed at the University of Wisconsin-Madison

Spraker, J., **K. A. Jewell**, L. V. Roze, J. Scherf, R. Beaudry, J. E. Linz, C. Allen, and N. Keller. **2014**. A volatile relationship: Profiling an inter-Kingdom dialogue between two plant pathogens, *Ralstonia solanacearum* and *Aspergillus flavus*. Journal of Chemical Ecology **40:**502-513.

de Oliveira, M. N. V., **K. A. Jewell**, F. S. Freitas, L. A. Benjamin, M. R. Tótola, A. C. Borges, C. A. Moraes, and G. Suen. **2013**. Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. Veterinary Microbiology **164:**307-314.

Christopherson, M. R., G. Suen, S. Bramhacharya, **K. A. Jewell**, F. O. Aylward, D. Mead, and P. J. Brumm. **2013**. The Genome Sequences of *Cellulomonas fimi* and "*Cellvibrio gilvus*" Reveal the Cellulolytic Strategies of Two Facultative Anaerobes, Transfer of "*Cellvibrio gilvus*" to the Genus *Cellulomonas*, and Proposal of *Cellulomonas gilvus* sp. nov. PLoS ONE **8:**e53954.

Chaston, J. M., G. Suen, S. L. Tucker, A. W. Andersen, A. Bhasin, E. Bode, H. B. Bode, A. O. Brachmann, C. E. Cowles, K. N. Cowles, C. Darby, L. de Léon, K. Drace, Z. Du, A. Givaudan, E. E. Herbert Tran, **K. A. Jewell**, J. J. Knack, K. C. Krasomil-Osterfeld, R. Kukor, A. Lanois, P. Latreille, N. K. Leimgruber, C. M. Lipke, R. Liu, X. Lu, E. C. Martens, P. R. Marri, C. Médigue, M. L. Menard, N. M. Miller, N. Morales-Soto, S. Norton, J.-C. Ogier, S. S. Orchard, D. Park, Y. Park, B. A. Qurollo, D. R. Sugar, G. R. Richards, Z. Rouy, B. Slominski, K. Slominski, H. Snyder, B. C. Tjaden, R. van der Hoeven, R. D. Welch, C. Wheeler, B. Xiang, B. Barbazuk, S. Gaudriault, B. Goodner, S. C. Slater, S. Forst, B. S. Goldman, and H. Goodrich-Blair. **2011**. The Entomopathogenic Bacterial Endosymbionts *Xenorhabdus* and *Photorhabdus*: Convergent Lifestyles from Divergent Genomes. PLoS ONE **6:**e27909.

Palmer, A. G., E. Streng, **K. A. Jewell**, and H. E. Blackwell. **2010**. Quorum sensing in bacterial species that use degenerate autoinducers can be tuned by using structurally identical non-native ligands. ChemBioChem **12:**138-147.